PARAMETERS OPTIMIZATION ALGORITHMS FOR IMPROVING THE PERFORMANCE OF OBSTACLES IDENTIFICATION IN FOREST AREA

/

提高林区障碍物识别效率的参数优化算法研究

Assoc. Prof. Ph.D. Lei Y., Ph.D. Xiaokang D., Ms. Jianlei K., Ms. Zheng Y., Prof. Ph.D. Jinhao L.
School of Technology,Beijing Forestry University / China
E-mail: liujinhao@vip.163.com

Abstract: *Accurate identification and measurement of obstacles in forest are particularly important for improving operational capability and efficiency of harvester. In this paper, three distinct algorithms, including grid search (GS), particle swarm optimization (PSO) and genetic algorithm (GA), are proposed for parameters optimization of support vector machine (SVM) to improve the performance of detecting obstacles in forest area. First, the three different optimization algorithms are respectively adopted to optimize parameters which are used in SVM. Then the SVMs with optimized parameters are applied to identify obstacles in forest. At last, their performances are compared based on the testing results of 150 samples of obstacles in the forest. The experimental results show that the GS algorithm with leave-one-out (LOO), PSO algorithm and GA algorithm have better classification accuracy than the GS with 6-fold cross-validation (CV). Especially for PSO and GA algorithms, the global optimal solution can be found rapidly without traversing all parameters within the grid.*

Keywords: *grid search,particle swarm optimization, genetic algorithm,support vector of machine, parameter optimization*

摘要: *准确的识别和测量林区障碍物，能够大大提高林业装备的工作性能和效率。本文提出了三种不同的参数优化算法来提高支持向量机（SVM）在林区障碍物识别中的作用效果，这三种算法分别是网格搜索（GS），粒子群优化（PSO）和遗传算法（GA）。首先，分别采用三种不同的优化算法来优化 SVM 中使用的参数。然后，将经过参数优化的支持向量机应用于林区障碍物的检测和识别。最后，基于 150 个样本的测试结果对不同算法的效果进行比较和分析。实验结果表明，利用留一法（LOO）的网格搜索，粒子群算法和遗传算法比利用 6 阶交叉验证的网格搜索具有更好的分类精度。尤其是 PSO 和 GA 算法，可以不用遍历网格内的所有参数而迅速找到全局最优解。*

关键词： *网格搜索，粒子群优化，遗传算法，支持向量机，参数优化*

INTRODUCTION

To meet the needs of forestry production, it is imperative to develop multi-function and high-efficiency forestry equipment to complete logging operation instead of manual operation. However, the complexity of forest environment will increase the operation risk and reduce the operating efficiency. Therefore, accurate identification of obstacle in forest and measurement of its feature are particularly important for improving operational capability and efficiency.

In respect of obstacle identification, researchers have done a lot of work and proposed a variety of algorithms.

Harls Baltzakls et al. proposed a suitably modified version of vector field histogram algorithm for robot motion planning and collision avoidance based on both laser and visual data [1]. Jaehyun Han et al. developed a road boundary and obstacle detection method using a downward-looking light detection and ranging sensor [2]. Yihui Lu et al. used Dempster-Shafer algorithm to automatically detect buildings from aerial images [3]. So far, obstacle detection method is mainly applied to vehicle systems, robots and building modeling. As regards to algorithms, Bayesian inference [4], [5], Dempster-Shafer evidence theory [6], [7], artificial neural net (ANN) [8], and support vector machine (SVM) classifier [9], [10] are popular and widely used in researches. Compared with other algorithms, SVM has many unique advantages in solving small sample, nonlinear and high dimensional pattern recognition [11]. However, like other machine learning algorithms, SVM performance depends highly on parameter selection. For a classification task, picking the best value for each variable is a model selection problem that needs an exhaustive search over the space of

研究背景

为了满足林业生产的需要，当务之急是要发展多功能、高效率的林业设备来代替手工操作完成采伐作业。然而，森林环境的复杂性会增加林业装备的操作风险，降低运行效率。因此，准确的识别和测量林区障碍对林业装备操作能力和效率的提高具有重要意义。

在障碍物识别方面，研究人员已经做了很多工作，并提出了各种各样的算法。

Harls Baltzakls 等，基于激光和可视化的数据对向量场直方图算法提出了修改，用于机器人的运动规划和避免碰撞 [1]。Jaehyun Han 等，利用下视角的光探测传感器提出了道路边界和障碍物检测方法 [2]。Yihui Lu 等使用 DS 证据算法自动检测建筑物航拍图像 [3]。到目前为止，障碍物检测方法主要适用于车辆系统，机器人和建筑建模。在算法方面，贝叶斯推理 [4], [5]，DS 证据推理理论 [6], [7]，人工神经网络（ANN）[8]，支持向量机（SVM）[9], [10]是近年来较受欢迎的算法并广泛应用于各种研究。与其他算法相比，SVM 在解决小样本，非线性及高维模式识别方面具有许多独特的优势 [11]。然而，像其他机器学习算法一样，支持向量机的分类性能高度依赖于参数的选择。对于分类任务来说，为每个参数变量选取最佳值需要详尽的搜索

hyper-parameters. Till now, most SVM practitioners selected these parameters only empirically by trying a finite number of values and retaining those with the least testing errors [12].

In parameter optimization, Grid search (GS) algorithm based on "Leave-One-Out" (LOO) is one widely used algorithm in machine learning. In this method, the parameter range of the approximate optimal value is firstly selected artificially, and then an exhaustive search on the set of parameters is conducted to obtain the optimal parameters [13]. Obviously, this method is time-consuming and the low efficiency is not desirable. In view of this, Chapelle et al. proposed a gradient descent method to automatically select SVM parameters [14], which made significant improvements in reducing computation time. However, it was often trapped in local optimal solution, and algorithm termination often occurred, which brought much trouble to the operation. Besides, Keerthi adopted quasi-Newton method for Gaussian kernel function parameter optimization of SVM model [15]. Particle swarm optimization (PSO) is a cluster optimization algorithm proposed by Kennedy and Eberhart in 1995 [16]. It is inspired by the movement behavior of birds and fish populations, and it is now the representative of swarm intelligence methods. Genetic algorithm (GA) was developed by Leung based on real-value to achieve automatic selection of the SVM model parameters [17]. This algorithm is used to select the parameters of SVM model based on the global optimal performance to improve the construction efficiency of SVM and the recognition rate of the classifier. During the past decades, SVM with GA and PSO have been applied in various fields, such as gene selection and classification [18], fault diagnosis [19], pattern recognition [20], and so on.

In this paper, the application field of parameter optimization method is extended to the forestry field for the first time and three different algorithms are proposed to identify obstacles by testing 150 obstacle samples in forest. Based on this, the principles and differences of the three algorithms are analyzed, and their advantages and disadvantages are summarized. The experimental results show that the GS algorithm with leave-one-out (LOO), PSO algorithm and GA algorithm have better classification accuracy than the GS with 6-fold cross-validation (CV).

PROPOSED PARAMETERS OPTIMIZATION METHOD

The parameters optimization process is shown in Figure 1, where three kinds of parameter optimization algorithms are listed as alternative methods. In this paper, RBF (Radius Basis Function) is selected as SVM kernel function and parameters that affect the performance of SVM are error penalty parameter C, and the kernel parameter σ. C represents the tolerable degree of the errors and σ is kernel width. Too high or too low value of the parameters will cause "over-learning" or "less-learning" in SVM so that it cannot identify samples effectively. In order to obtain appropriate C and σ, three algorithms are adopted to optimize the parameters. Next, the principle and flow of each algorithm will be illustrated for better understanding of testing results.

超参数的空间，这是一个复杂的模型选择问题。目前，大多数的 SVM 研究者选择参数是通过经验尝试有限数量的值，然后选择能够产生最小测试误差的参数 [12]。

在参数优化中，基于留一法（LOO）的网格搜索（GS）算法是一种被广泛使用的机器学习算法。在该方法中，首先在近似最优值的参数范围内进行人工选择，然后在参数集中进行穷举搜索以获取最佳的参数值 [13]。显然，这种方法是费时的，并且效率低。因此，Chapelle 等提出了一种梯度下降法来自动选择 SVM 参数 [14]，在减少计算时间方面有显着改善。然而，该算法容易陷入局部最优解，算法终止的经常发生为实际操作带来了很多麻烦。除此以外，Keerthi 采用了拟牛顿法高斯核函数参数优化算法来优化 SVM 模型 [15]。另外，粒子群优化（PSO）是由 Kennedy 和 Eberhart 于 1995 年提出一个集群优化算法 [16]。它的灵感来自于鸟类和鱼类的数量流动行为，现在是群体智能的代表算法。遗传算法（GA）是由 Leung 提出的 SVM 模型参数自动选择方法 [17]。该算法基于对全球最佳的性能来选取 SVM 参数，提高了 SVM 的操作效率和分类识别率。在过去的几十年里，基于 GA 和 PSO 的 SVM 已经应用在各个领域，如基因的筛选和分级 [18]，故障诊断 [19]，模式识别 [20]，等等。

在本文中，参数优化方法的应用领域第一次扩展到了林区，提出了三种不同的算法来识别林区障碍，通过测试 150 个林区障碍样品，分析了三种算法的原理和效果，并对它们的优点和缺点进行总结。实验结果表明，利用留一法（LOO）的网格搜索，粒子群算法和遗传算法比利用 6 阶交叉验证的网格搜索具有更好的分类精度。

参数优化算法

整个参数优化过程如图 1 所示，三种不同优化算法分别对 SVM 模型进行参数选择，用优化后的 SVM 模型对林区障碍物进行识别并输出结果。本文中 SVM 模型的核函数选择的是最常用的径向基函数 RBF(Radius Basis Function)，因此影响 SVM 模型性能的参数有误差惩罚参数 C 和核参数 σ。C 代表错误的容忍程度，σ 是核宽度。过高或过低的参数值会造成 SVM 模型的"过学习"或"学习不足"现象，造成 SVM 不能有效对目标进行分类。为了获得适当的 C 和 σ 值，采用了三种算法对参数进行优化。下面，对每个算法的原理和流程进行说明便于更好地理解测试结果。

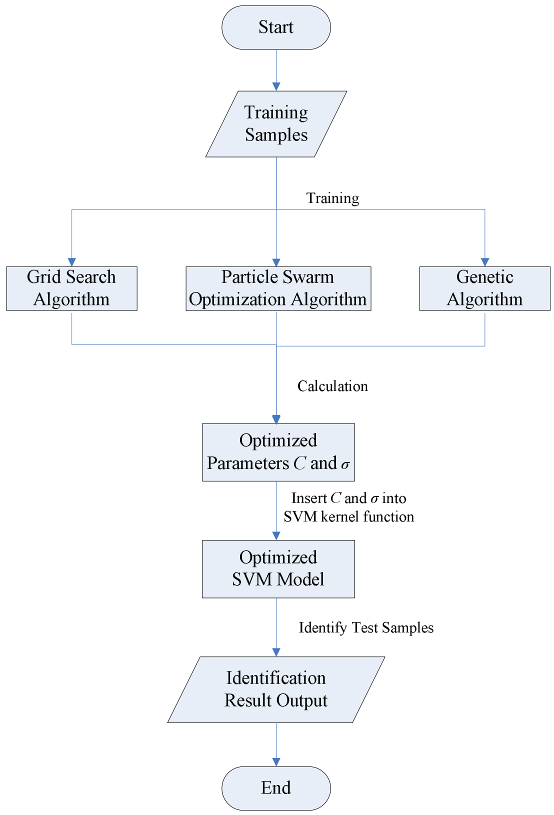

Fig. 1 - Flowchart of the parameters optimization process / *参数优化算法流程图*

Grid Search Algorithm

The basic principle of GS algorithm is that divide C and σ into grids within a certain range, and traverse all points within the grid values, then use the cross-validation (CV) method to validate each value of C and σ, finally C and σ with the best classification accuracy is chosen as the optimal parameters.

The CV method is a statistical analysis method which is used to validate the performance of classifier. The algorithm flow of k-fold CV is as follows:

(1) Initialize the parameters C and σ of SVM classifier;

(2) Randomly divide the training data into k mutually exclusive subsets of approximately equal size, and use one subset as testing set and the other k-1 subsets as training set to evaluate classifier performance;

(3) Calculate the classification accuracy with the initialized parameters, and repeat this procedure k times to ensure that each subset is used once for testing;

(4) Choose the parameters with the highest classification accuracy.

Leave-one-out (LOO) can be viewed as an extreme form of k-fold cross-validation in which k is equal to the number of examples.

网格搜索算法

GS 算法的基本原理是：将 *C* 和*σ*的取值在一定范围内划分为网格，遍历所有网格内的值，利用交叉验证法 CV（Cross-Validation）来验证每一个 *C* 和*σ*的值，最终选取具有最佳分类精度的 *C* 和*σ*作为最优参数。

CV 方法是一种统计分析方法，用来验证分类器的性能。K-阶 CV 算法流程如下：

(1) 初始化 SVM 分类器参数 *C* 和*σ*；

(2) 随机将训练数据分成 *k* 份互斥且大小相等的子集，并使用其中一个子集作为测试集而其他 *k-1* 个子集作为训练集来检测分类器的性能；

(3) 计算经过参数初始化的 SVM 的分类精度，并重复此过程 k 次以确保每个子集被作为一次测试集；

(4) 选择具有最高分类精度的参数值作为最佳参数。

留一法（LOO）可以被看作是一种极端形式的 k-阶交叉验证法，其中，k 等于被测样本的数量。

PSO Algorithm

PSO is a new heuristic global search algorithm based on swarm intelligence, and it performs the global optimum search via competition and collaboration between the particles in a complex search space. In PSO algorithm, if there is a particle swarm composed of N particles in D-dimensional space, then each particle is a possible solution in the D-dimensional search space. Particles are moving in solution space, and dynamically adjust the speed and direction according to the instantaneous optimal solution of each particle and the whole population.

粒子群算法

PSO 是基于群体智能的一种新的启发式全局搜索算法，它通过在复杂的搜索空间粒子之间的竞争和协作实现全局最优搜索。在 PSO 算法中，如果在一个 D 维空间中有 N 个粒子组成的粒子群，每个粒子是一个在 D 维搜索空间中可能的解。粒子在解空间移动，并根据每个粒子和整个群体的瞬时最优解动态地调整各自的速度和方向。

In a D-dimensional search space, a community is composed of N particles, of which the i-th particle is represented by a D-dimensional vector:

在一个 D 维搜索空间，一个群体是由 N 个粒子组成，其中第 i 个粒子可以由一个 D 维向量表示如下：

$$X_i = (x_{i1}, x_{i2}, \cdots, x_{iD}), i = 1,2,\cdots,N \tag{1}$$

The "flight speed" of the i-th particle is also a D-dimensional vector, denoted by:

第 i 个粒子的飞行速度也是一个 D 维的向量，表示为：

$$V_i = (v_{i1}, v_{i2}, \cdots, v_{iD}),\ i = 1,2,\cdots,N \tag{2}$$

The instantaneous best position of a particle i is called individual extremum, denoted by:

粒子瞬时最佳位置 i 被称为个体极值，记为：

$$p_{best} = (p_{i1}, p_{i2}, \cdots, p_{iD}),\ i = 1,2,\cdots,N \tag{3}$$

The instantaneous best position of the whole particle swarm is called global extremum, denoted by:

整个粒子群的瞬时最佳位置被称为全局极值，记为：第 i 个粒子的飞行速度也是一个 D 维的向量，表示为：

$$g_{best} = (p_{g1}, p_{g2}, \cdots, p_{gD}) \tag{4}$$

In the searching process, particles adjust the speed and direction according to the following formulae:

在搜索过程中，粒子根据下列公式调整速度和方向：

$$v_{id}(t+1) = w * v_{id}(t) + c_1 r_1 (p_{id}(t) - x_{id}(t)) + c_2 r_2 (p_{gd}(t) - x_{id}(t)) \tag{5}$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \tag{6}$$

where c1, c2 are acceleration constant, r1 and r2 are the uniform random value in the range of [0, 1] .

The algorithm flow of PSO is as follows:

(1) Initialize the swarm, including population size N, the position x_i and velocity V_i of each particle;

(2) Calculate the fitness value $F_{it}[i]$ of each particle;

(3) For each particle, compare its fitness value $F_{it}[i]$ with individual extremum value $p_{best}(i)$. If $F_{it}[i] > p_{best}(i)$, replace $p_{best}(i)$ by $F_{it}[i]$;

(4) For each particle, compare its fitness value $F_{it}[i]$ with global extremum value g_{best}. If $F_{it}[i] > g_{best}$, replace g_{best} by $F_{it}[i]$;

(5) Update the particle velocity v_i and position x_i according to the equation (1) and (2);

(6) Output the result when the error is low enough or the maximum number of iteration is reached; otherwise return to step (2).

其中，c1，c2 为加速常数，r1 和 r2 是在[0, 1]范围内的随机值。PSO 算法流程如下：

（1）初始化粒子群，包括群体规模 N，每个粒子的位置 x_i 和速度 V_i；

（2）计算每个粒子的适应值 $F_{it}[i]$；

（3）对于每个粒子，比较它的适应值 $F_{it}[i]$ 与个体极值 $p_{best}(i)$。如果 $F_{it}[i] > p_{best}(i)$，用 $F_{it}[i]$ 替代 $p_{best}(i)$；

（4）对于每个粒子，比较它的适应值 $F_{it}[i]$ 与全局极值 g_{best}。如果 $F_{it}[i] > g_{best}$，用 $F_{it}[i]$ 替代 g_{best}；

（5）根据公式(1)和(2)更新粒子的速度和位置；

（6）当误差足够低或最大迭代次数达到最大值时输出结果，否则返回到步骤（2）。

GA Algorithm

The searching for the optimal solution of GA is an imitation process of biological evolution, which is done by chromosome crossover and mutation. The GA algorithm mainly uses the selection operator, crossover operator and mutation operator to simulate biological evolution, and produce generation after generation of the population. The three operations are defined according to evolutionary terms as follows. Selection operator is used to select individuals that adapt to the environment from the group by calculating fitness value, and these individuals are selected for breeding next generation. For the selected individuals, crossover operator is used to exchange genes in the same position of two different individuals based on some crossover probability, and mutation operator is used to change genes of some individuals, which is a simulation of gene mutation.

In GA, the initial solution group is composed of n binary string of length L. In each string, each binary bit is its individual chromosome gene. In a binary string, if a binary bit is 1, then it will be converted to 0 after mutation operation, and vice versa. The algorithm flow of GA is as follows:

遗传算法

GA 是一种模仿生物进化过程的算法，通过染色体的交叉和变异完成。遗传算法主要采用选择算子，交叉算子和变异算子模拟生物进化，并产生后代。三个算子操作定义根据进化条款如下：选择算子是通过计算适应值来选择适应环境的个体，然后选择这些个体来繁殖下一代。对于选定的体，交叉算子根据一定的交叉概率交换在同一位置的两个不同个体的基因，变异算子通过改变某些个体的基因来模拟基因突变的过程。

在遗传算法中，初始解组是由 n 个长度为 L 的二进制串组成。在每个字符串中，每个二进制位是个体的染色体基因。在一个二进制串中，如果一个二进制位是 1，经过变异操作后它会被转换为 0，反之亦然。

遗传算法的算法流程如下：

(1) Initialize the population, including population size N, crossover probability P_c, mutation probability P_m and the standard of the evolutionary termination;
(2) Calculate the fitness value of each individuals;
(3) Achieve population evolution by selection operator, crossover operator and mutation operator;
(4) Output the result when the standard of evolution termination is met; otherwise return to step (2).

（1）初始化的粒子群，包括群体规模 N，交叉概率 P_c，变异概率 P_m 和进化终止的条件；
（2）计算每个个体的适应值；
（3）通过选择算子，交叉算子和变异算子实现种群进化；
（4）当满足进化终止条件时输出结果，否则返回步骤（2）。

Support Vector Machine

In this paper, SVM model is used to identify obstacles in forest. As a new machine learning algorithm, its core idea is to convert nonlinear separable problems in low-dimensional space into linearly-separable problems in high-dimensional space.

If there are two classes in space H can be separated by a hyperplane as follows:

支持向量机 SVM

在本文中，SVM 模型被用来识别林区的障碍物。作为一种新的机器学习算法，其核心思想是将低维空间中非线性可分问题转换成在高维空间中线性可分问题。如果有两个类可以在空间 H 中由超平面进行分类，那么超平面表示如下：

$$\mathrm{w} \cdot x + \mathrm{b} = 0 \tag{7}$$

where x represents the feature vector, our goal is to calculate the values of w and b to determine the optimal hyperplane which maximizes the margins of the two classes (Figure 2).

其中 x 表示特征向量，我们的目标是计算 w 和 b 的值，以确定得到最佳的超平面，能够实现这两个类的间距最大化（图 2）。

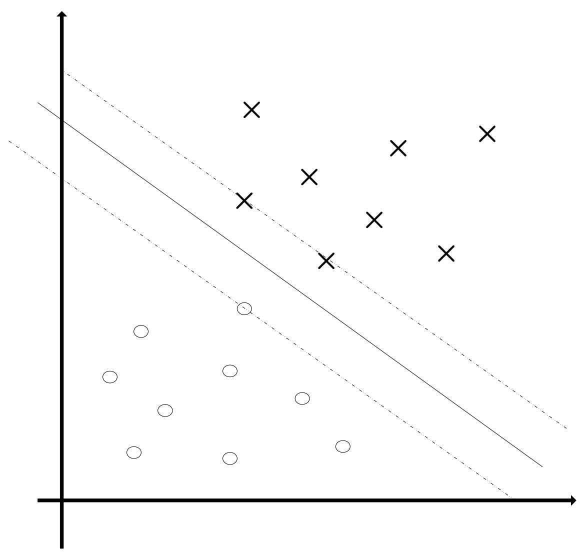

Fig. 2 - SVM maximum interval hyperplane / *SVM 最大间隔超平面*

In SVM model, it is not necessary to make all vectors far away from the hyperplane. What we care about most are the vectors that are nearest to hyperplane. By calculating the distance between the nearest vectors to the optimal hyperplane, the question can be described by the following formula:

在 SVM 模型中，没有必要使所有的向量远离超平面，而需要关心的是距离超平面最近的向量。通过计算距离最优超平面最近的向量之间的距离，问题可以由下式描述：

$$\min_{w,b} \frac{1}{2}\|w\|^2$$
$$s.t. y^{(i)}(w^T x^{(i)} + b) \geq 1, i = 1, 2, ..., m \tag{8}$$

To solve the above problem, Lagrange operator is introduced to the formula. By simplification and conversion, the optimal values of w and b can be obtained by solving a constrained minimization problem as follows:

为了解决上述问题，拉格朗日算子被引入到公式中。通过简化和转换，解决如下约束最小化问题，可以获得 w 和 b 的最佳值：

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b\right) \tag{9}$$

where *K (xi, x)* is the kernel function used to solve nonlinear separable problem. Commonly used kernel functions are linear kernel, polynomial kernel, RBF (Radius Basis Function) kernel and so on. In this paper, the SVM model parameter selection issue is considered based on RBF kernel as follows:

其中 K(x_i, x)是用于解决非线性可分问题的核函数。常用的核函数有线性核函数，多项式核函数，径向基函数（RBF）核函数等。在本文中，SVM 模型参数选择问题被认为是基于 RBF 核函数的，该核函数表示如下：

$$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \tag{10}$$

Therefore, the performance of the SVM model depends on the error penalty parameter *C* and the kernel parameter *σ*. The above three algorithms are used to optimize the value of parameter *C* and *σ*. After that, the kernel function with optimized *C* and *σ* is adopted to construct SVM model to improve the obstacles identification performance.

因此，SVM 模型的性能取决于误差惩罚参数 *C* 和核参数 *σ*。上述三个算法用来优化的参数 *C* 和*σ*的值。然后，将优化后的 *C* 和*σ*代入 SVM 模型，以改善模型对林区障碍物的识别性能。

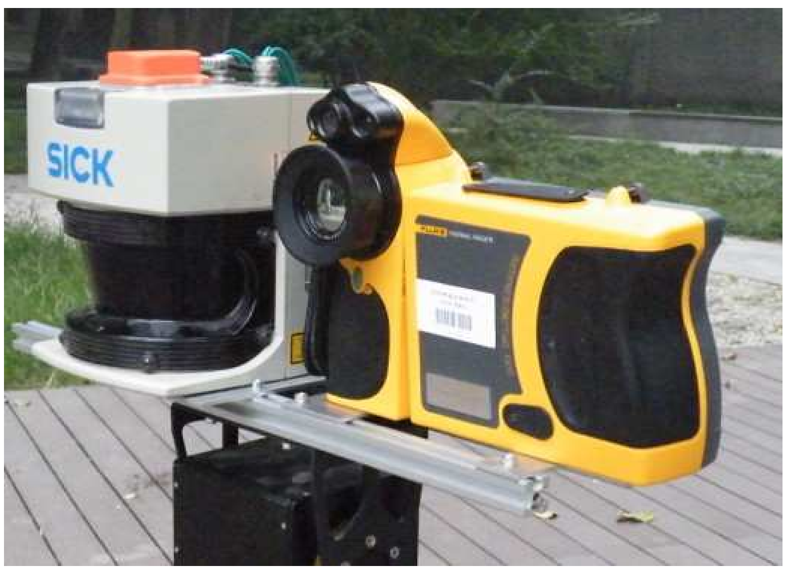

Fig. 3 - Sensors placed on the same horizontal plane in this experiment / *试验设备与装置*

EXPERIMENTAL RESULTS AND DISCUSSION

Figure 3 shows the 2D laser scanner and infrared thermal imager used for detecting obstacles in this experiment, and the sensors of the scanner and imager are placed on the same horizontal plane. Firstly, the position and image of obstacle is collected by the equipment. Then, after images fusion and data association, the features of the objects such as the height, width, shape, temperature, and color could be obtained from the laser points, visible images, and infrared images. Finally, three SVM parameter optimization algorithms, of which the GS algorithm includes 6-fold crossover-validate and LOO, are applied to a RBF SVM model to achieve obstacles recognition. In order to examine and compare the SVM performance, the three algorithms are tested using 150 forest obstacles samples, including 50 trees, 50 people, and 50 stones. For each type of obstacle, 40 samples were selected randomly as the training data and the other 10 as the test data (Table 1).

试验结果与讨论

图 3 显示了在本实验中用于检测林区障碍物的 2D 激光扫描仪和红外热像仪，激光扫描仪和红外热像仪被装置在同一水平面上。首先，通过两种传感器可以获得障碍物的位置和图像信息。然后，经过图像融合和数据关联，通过获得的激光点，可见光图像和红外图像可以得到被测对象的特征，如高度，宽度，形状，温度，颜色等。最后，三种 SVM 参数优化算法，其中 GS 算法中分别使用了 6 阶交叉验证和留一法，被应用于基于 RBF 的 SVM 模型来实现林区障碍识别。为了检验和比较支持向量机的性能，使用 150 个森林障碍物的样本（包括 50 棵树，50 人，50 石头）来进行结果测试。对于每一类型的障碍物，随机选取 40 个样本作为训练数据，剩余 10 个样本作为测试数据（表 1）。

Table 1 / *表 1*

Training data and test data / *训练数据和测试数据*

Sample Type / 样本类型	Total Samples / 样本总数	Training Samples / 训练样本	Test Samples / 测试样本
Tree / 树木	50	40	10
People / 人	50	40	10
Stone / 石头	50	40	10
Total / 总数	150	120	30

In GS algorithm, firstly the range and search step of the variables are set, where $C\in$ $[2^{-8}, 2^{8}]$, $\sigma\in$ $[2^{-8}, 2^{8}]$ and the search step is 0.5 for both C and σ. Then 6-fold CV and LOO methods are adopted respectively to validate the value of C and σ. Based on the calculation, when k > 6, the growth of the best accuracy by CV is relative slow, so the k is set as 6. Moreover, LOO method is used as an extreme of CV to be compared with 6-fold CV.

In PSO, acceleration constant c_1=c_2=2, inertia weight ω=1, $C\in$ (0, 100], $\sigma\in$ [0, 1000]. Population size is set to 20 and the maximum iteration generation is 200. When the iteration time is up to 200 or the global fitness remains unchanged in 100 consecutive iterations, the operation stops.

In GA, acceleration constant c_1=c_2=2, inertia weight ω=1, $C\in$ (0, 100], $\sigma\in$ [0, 1000]. Population size is set to 20 and the maximum iteration generation is 200. When the iteration time is up to 200 or the global fitness remains unchanged in 100 consecutive iterations, the operation stops.

After parameter settings, SVM model parameters can be calculated by different algorithms based on 120 training samples, and Table 2 presents one group of calculation results as an example.

下面对不同算法的参数进行设置：在 GS 算法中，首先设置变量的范围和搜索步骤，其中 C∈$[2^{-8}, 2^{8}]$，σ∈$[2^{-8}, 2^{8}]$，搜索步长均为 0.5。通过计算可知，识别精度随着 k 值的增长而增长，而当 k> 6 时，识别精度的增长较慢，所以 k 被设置为 6。此外，LOO 法作为一个极端例子与 6 阶 CV 进行比较。

在 PSO 算法中，加速度常数 c_1 = c_2= 2，惯性权重ω=1，C∈（0，100]，σ∈[0，1000]。群体大小设置为 20，最大迭代次数为 200。当迭代次数高于 200 或全局极值连续迭代 100 次不变时，停止操作。

在遗传算法中，加速度常数 c_1 = c_2=2，惯性权重ω=1，C∈（0，100]，σ∈[0，1000]。群体大小设置为 20，最大迭代次数为 200。当迭代次数高于 200 或全局极值连续迭代 100 次不变时，停止操作。

经过以上参数设置，同时基于 120 个训练样本的测试，可以获得不同优化算法优化后的 SVM 模型参数，如表 2 所示为一组计算结果。

Table 2 / *表 2*

SVM model parameters obtained with different algorithms / *经过不同优化算法优化后的参数*

Algorithm / *算法*		Best Fitness / *最佳适应度*	Best C / *最佳 C 值*	Best σ / *最佳 σ 值*
GS / *网格搜索*	6-fold / *6 阶*	92.50%	2	0.0313
	LOO / *利用留一法*	95.00%	2	0.0039
PSO / *粒子群优化*		94.17%	3	0.0100
GA / *遗传算法*		95.83%	2.89	0.0019

Due to the parameters obtained from the three algorithms, the SVM classifier can be used to classify the test samples, and their performances are shown in Figure3, according to which the average classification accuracy of each parameter algorithm can be calculated.

根据三种算法获得的参数，SVM 分类器可以用于分类测试样本，它们的测试结果如图 3 所示，由测试结果可以计算出每个算法的平均分类精度。

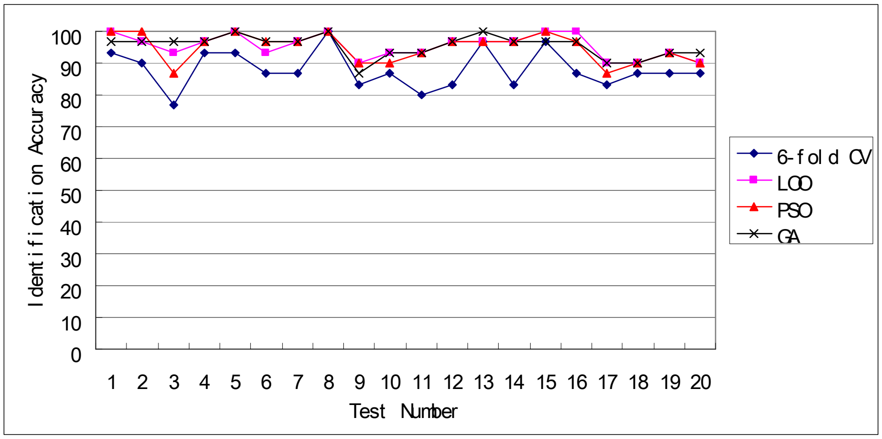

Fig. 3 - Identification accuracy of each algorithm in 20 tests / *三种算法的障碍物识别准确度结果*

The average recognition accuracy of GS with 6-fold CV, GS with LOO, PSO and GA are respectively 91%, 95.33%, 94.85% and 95.34%. It is obvious that the SVM model with the 6-fold CV has relative lower classification accuracy compared with the other three algorithms. In GS algorithm, almost all samples are used to train the model in each iteration for LOO method, which is the closest to the original sample distribution, so that LOO method has a better effect than 6-fold CV method dose, and the assessment outcome of the LOO method is more reliable. However, LOO method has the shortcoming of high computing cost, because the number of models requiring to be established is the same as the original number of data samples.

The identification accuracy of GS algorithm with LOO method and GA algorithm both can reach over than 95%, which is slightly better than that of PSO algorithm. GA algorithm and PSO algorithm are both trying to simulate the population adaptability on the basis of natural characteristics, and they adopt certain transformation rules to solve problems by searching space. In our study, these two algorithms both have high classification accuracy for obstacle detection in the forest. They have their own characteristics and advantages, as well as defects and deficiencies. PSO algorithm has more efficient information sharing mechanism than GA algorithm and all particles in PSO algorithm may converge faster to the optimal solution than the evolutionary individuals in GA algorithm. But this mechanism may lead to over-concentration of particles, which is likely to fall into the local minimum. For GA algorithm, coding techniques and genetic manipulation are simple, while PSO algorithm has no codes or crossover and mutation operations, and the particles are updated only by the internal speed. Therefore, the principle of PSO algorithm is more simply and easily to achieve.

Besides those above, it should be noted that although GS algorithm, especially with LOO method, can find the global optimal solution, it will be very time-consuming sometimes if people want to find the best parameters C and σ in a larger range. But for GA and PSO algorithms, they are both heuristic algorithms that can find the global optimal solution without traversing all parameters within the grid, which will save a lot of computing cost.

CONCLUSIONS

Three parameters optimization algorithms: GS algorithm, PSO algorithm and GA algorithm were proposed to a SVM classifier to identify obstacles in forest area in this paper. The principle and flow of each algorithm were introduced and the parameters of each algorithm were set according to experience. After that, parameters C

6 阶 CV，LOO，PSO 和 GA 的平均识别精度分别为 91%，95.33%，94.85%和 95.34%。显然，6 阶 CV 的 SVM 模型与其他算法相比，具有较低的分类精度。在 LOO 算法中，每次迭代几乎用到所有的样本来训练模型，这是最接近原始样本分布的方法，所以 LOO CV 法比 6 阶 CV 法具有较高的准确度，评估结果显示 LOO 方法更可靠。然而，因为需要建立的模型数目与原始的数据样本数量是相同的，所以 LOO 方法具有计算成本高的缺点。

LOO 法和遗传算法的识别准确度都可以达到 95%以上，略优于 PSO 算法。GA 算法和 PSO 算法都试图模拟基于自然特性的种群适应性，并且采取一定的变换规则，通过搜索空间来解决问题。在我们的研究中，这两种算法在森林的障碍物检测中都具有较高的分类准确率。它们有自己的特点优势，然而也有缺陷和不足。PSO 算法比遗传算法具有更高效的信息共享机制，并且 PSO 算法中的所有粒子比遗传算法的进化个体更快收敛于最佳解。但这种机制可能会导致粒子过度集中而陷入局部最小值。GA 算法的编码技术和遗传操纵方法虽然简单，而 PSO 算法不需要任何编码或交叉和变异操作，粒子仅进行内部速度更新。因此，PSO 算法的原理更简单，更容易实现。

除了上述分析结果，应当指出的是，虽然 GS 算法可以找到全局最优解，但是如果想在一个较大的范围内找到最佳 C 和σ是非常费时的。而 GA 和 PSO 算法都是启发式算法，可以在网格内无需遍历所有参数找到全局最优解，节省了大量的计算成本。

结论

文中提出了三种参数优化算法：GS 算法，粒子群算法和遗传算法，并将它们用于 SVM 分类器来对林区障碍物进行识别。介绍了每种算法的原理和流程，并根据经验设置了相应参数。然后，运用这些优化算法对 SVM 模型中的参数

and σ that were used in SVM were optimized by these algorithms. Then the optimized C and σ were inserted into kernel function in SVM to identify three kinds of obstacles in forest. Finally, the performances of the different algorithms were examined and the experimental results showed that the GS algorithm with LOO, PSO algorithm and GA algorithm had better classification accuracy than GS algorithm with 6-fold CV. Considering the time-consuming disadvantage of GS algorithm with LOO, the PSO and GA algorithm are considered as more suitable methods for obstacle recognition in the forest. Our work extends the application field of SVM to forestry, which may provide valuable reference for developing new forestry equipment.

Aknowlegement

This study is financially supported by National Natural Science Foundation of China (Grant No.31070634), China Postdoctoral Science Foundation (Grant No. 2012M510330), 948 project supported by State Forestry Administration, China (Grant No. 2011-4-02).

REFERENCES

[1]. Baltzakis H., Argyros A., Trahanias P., (2003) - "*Fusion of laser and visual data for robot motion planning and collision avoidance*," Machine Vision and Applications, 92-100,15(2);
[2]. Han J.; Kim D.; Lee M.; Sunwoo M., (2012) - "*Enhanced road boundary and obstacle detection using a downward-looking lidar sensor*," Ieee Transactions on Vehicular Technology, 61;
[3]. Lu Y. H., Trinder J. C., Kubik K., (2006) - "*Automatic building detection using the Dempster-Shafer algorithm," Photogrammetric engineering and remote sensing*, 395-403, 72 (4);
[4]. Pizarro, D.; Mazo, M.; Santiso, E.; Marron, M.; Jimenez, D.; Cobreces, S.; Losada, C., (2010) - "*Localization of mobile robots using odometry and an external vision sensor*," Sensors ,3655-3680, 10;
[5]. Zhu, Y.D.; Fujimura, K., (2010) - "*A bayesian framework for human body pose tracking from depth image sequences*," Sensors, 5280-5293, 10;
[6]. Khoshelham K.; Nardinocchi C., (2009) - "*A learning Dempster-Shafer model for automated building detection*," In Proceedings of the Optical 3-D measurement techniques IX, Vienna, Austria, 1-3, July;
[7]. Rottensteiner F.; Trinder J.; Clode S.; Kubik K.; Lovell B. (2004) - "*Building detection by Dempster-Shafer fusion of LIDAR data and multispectral aerial imagery*," In Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), 23-26, Aug.;
[8]. Shutao L., James T. K., Yaonan W., (2002) - "*Multifocus image fusion using artificial neural networks*," Pattern Recognition Letters, Vol.23, No.8, 985–997, June;
[9]. Bo W., Jixiang L., Qixiang Y., Zhenjun H., Jianbin J., (2011) - "*Fast pedestrian detection with laser and image data fusion*," 2011 Sixth International Conference on Image and Graphics (ICIG), 12-15, Aug.;
[10]. M. A. Sotelo., I. Parra., D. Fern´andez., E. Naranjo., (2006) - "*Pedestrian Detection using SVM and Multi-feature Combination*," In Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference, Toronto, Canada, 17-20, September;
[11]. V. Vapnik, (1995) - "*The Nature of Statistical Learning Theory*", Springer, NY;
[12]. Ayat N E, Cheriet M, Suen C Y., (2005) - "*Automatic model selection for the optimization of SVM kernels*," Pattern Recognition, 1733-1745, 38(10);
[13]. Lee J. H., Lin C. J., (2000) - *Automatic Model Selection for Support Vector Machines [R].* Taipei:Taiwan University;

C 和σ进行了优化。将优化后的 C 和σ分别代入 SVM 模型中，对林区中的三种常见障碍物进行识别和分类。最后，用 150 组样本数据对三种算法的识别结果进行了测试，实验结果表明，利用留一法（LOO）的网格搜索，粒子群算法和遗传算法比利用 6 阶交叉验证的网格搜索具有更好的分类精度。考虑到 LOO 法计算成本高的缺点，粒子群算法和遗传算法更适用于林区的障碍物识别。本文将支持向量机的应用扩展到了林业领域，为新型林业设备的开发提供了有价值的参考。

致谢

本项研究是由国家自然科学基金（批准 No.31070634），中国博士后科学基金（批准号：2012M510330），中国国家林业局 948 项目（批准号：2011-4-02）财政支持。

参考文献

[1]. Baltzakis H, Argyros A, Trahanias P., (2003) - *基于激光和视觉数据融合的机器人运动规划和避免碰撞研究*，计算机视觉与应用, 92-100, 15 (2);
[2]. Han J.; Kim D.; Lee M.; Sunwoo M., (2012) - *基于下视角激光雷达传感器的增强道路边界和障碍检测研究*，Ieee 车辆技术会议，61;
[3]. Lu Y. H., Trinder J. C., Kubik K., (2006) - *利用DS 算法自动检测建筑物*， 摄影测量工程与遥感 ，395-403, 72 (4);
[4]. Pizarro D.; Mazo M.; Santiso E.; Marron M.; Jimenez D.; Cobreces S.; Losada C., (2010) - *基于测程法和外部视觉传感器的移动机器人定位*， 传感器，3655-3680, 10;
[5]. Zhu Y.D.; Fujimura K., (2010) - 利用贝叶斯框架从深度图像序列跟踪人体姿态，传感器 ，5280-5293, 10;
[6]. Khoshelham K.; Nardinocchi C., (2009) - *利用学习的D-S模型自动检测建筑物*， 光学三维测量技术会议, Vienna, Austria，1-3, July;
[7]. Rottensteiner F.; Trinder J.; Clode S.; Kubik K.; Lovell B., (2004) - *基于激光雷达数据和多光谱航空图像DS融合的建筑检测*, 第17届模式识别国际会议 (ICPR'04), 23-26, Aug.;
[8]. Shutao L., James T. K., Yaonan W., (2002) - *基于人工神经网络的多焦点图像融合*，模式识别快报，Vol.23, No. 8, 985–997, June;
[9]. Bo W., Jixiang L., Qixiang Y., Zhenjun H., Jianbin J., (2011) - *吴波，梁吉祥， 叶启祥，韩哲君， 焦建斌；基于激光和图像数据融合的快速行人检测*， 2011 年第六届图像图形学国际会议 (ICIG),12-15, Aug.;
[10]. M. A. Sotelo., I. Parra., D. Fern´andez., E. Naranjo., (2006) - *基于SVM和多特征融合的行人检测方法研究*，2006 IEEE 智能交通系统会议, Toronto, Canada, 17-20, September;
[11]. V. Vapnik, (1995) - *统计学习理论的本质*，Springer, NY;
[12]. Ayat N E, Cheriet M, Suen C Y., (2005) - *SVM核函数优化的自动模型选择*，模式识别，1733-1745, 38 (10);
[13]. Lee J. H., Lin C. J., (2000) - *支持向量机自动模型选择*，Taipei: Taiwan University;

[14]. Chapelle O., Vapnik V., Bousquet O., et al. (2002) - "*Choosing Multiple Parameters for Support Vector Machines," Machine Learning*, 131-159, 46 (1-3);

[15]. Keerthi S. S. (2002) - "*Efficient tuning of SVM hyperparameters using radius / margin bound and iterative algorithms*," Neural Networks, IEEE Transactions on, 1225-1229, 13 (5);

[16]. Kennedy J., Eberhart R., (1995) - "*Particle swarm optimization," Neural Networks*, 1942-1948, 4;

[17]. Leung Y. W., Wang Y., (2001) - "*An orthogonal genetic algorithm with quantization for global numerical optimization*," IEEE Trans Evolutionary Computation, 41-53,5 (1);

[18]. Huerta E. B., Duval B., Hao J. K., (2006) - "*A hybrid GA/SVM approach for gene selection and classification of microarray data*," Applications of Evolutionary Computing. Springer Berlin Heidelberg, 34-44;

[19]. Wang J., LU S., (2005) - "*The study of GA-SVM on turbo generator fault diagnosis*," Turbine Technology, 10-20;

[20]. Ranaee V., Ebrahimzadeh A., Ghaderi R., (2010) - "*Application of the PSO-SVM model for recognition of control chart patterns*," ISA transactions, 577.49 (4).

[14]. Chapelle O., Vapnik V., Bousquet O., et al. (2002) - *支持向量机的多参数选择，机器学习*, 131-159, 46 (1-3);

[15]. Keerthi S. S. (2002) - *基于半径/边缘范围和迭代算法高效优化支持向量机超参数*，神经网络, 1225-1229,13 (5);

[16]. Kennedy J., Eberhart R., (1995) - *粒子群算法*， 神经网络, 1942-1948, 4;

[17]. Leung Y. W., Wang Y., (2001) - *全球数值优化的量化正交遗传算法*， IEEE 进化计算.41-53, 5 (1);

[18]. Huerta E. B., Duval B., Hao J. K., (2006) - *一个混合GA / SVM方法对基因微阵列数据的选择和分类*， 进化计算应用. Springer Berlin Heidelberg, 34-44;

[19]. Wang J., LU S., (2005) - *基于遗传算法支持向量机的汽轮发电机故障诊断研究*， 涡轮技术, 10-20;

[20]. Ranaee V., Ebrahimzadeh A., Ghaderi R., (2010) - *应用粒子群优化算法的支持向量机模型识别控制图模式*， ISA 学报, 577.49 (4).