Optimization and Application of Support Vector Machine Based on SVM Algorithm Parameters

YAN Hui-feng[1], WANG Wei-feng[1], LIU Jie[2]
[1]ChongQing University of Posts and Telecom
400065, China
[2]School Of Civil Engineering & Architecture
Chongqing Jiaotong Univ.,40074,China
4423360@qq.com, liu_jie@163.com

JDIM
Journal of Digital Information Management

ABSTRACT: *The hospital customer classification is very important for the integration and allocation of hospital resources, can greatly enhance the market competitiveness of the hospital. Support Vector Machine (SVM) is an approach to solve classification problem by using optimization method. Selecting different kernel parameters can construct different classifiers, meanwhile parameters decide their learning and generalization ability. In order to solve the limitation of selecting parameters by experience, so the particle swarm (PSO) pattern search algorithm is proposed to search optimal parameters and take them into the practice of hospital customer classification; The PSO mode search algorithm is to combine the advantages of PSO algorithm and pattern search algorithm, PSO mode search algorithm has strong global search capability and good advantage of local convergence. The result of experiment shows that this method is not only efficient, but also to search the optimal parameters achieving a high accuracy, which is an effective method of SVM parameter optimization.*

Categories and Subject Descriptors:
I.1.2: [Theory and algorithms]: Support vector machines; **E.1: [Database theory]** Data structures and algorithms for data management

General Terms:
Data Structures, Particle Swarm Optimization, Support Vector Machines

Keywords: Support Vector Machine, Kernel Parameters Selection, Particle Swarm Pattern Search, Customer Data Classification

Received: 1 November 2012, Revised 27 December 2012, Accepted 30 December 2012

1. Introduction

Support vector machines (Support Vector Machine, SVM) is Corinna Cortes and Vapnik, equal to the 1995 first proposed it in resolving the small samples, nonlinear and high dimensional pattern recognition exhibit many unique advantages, and be able to promote the application to function fitting and other machine learning problems[1]. In addition, support vector machines is a convex quadratic programming, extreme value is obtained to ensure the global optimal solution. In recent years, support vector machines in text categorization, remote sensing image analysis, fault identification and prediction, time series forecasting, and other fields have a successful application[2,3]. Support vector machine (SVM) classification, the choice of the kernel function is very important, kernel function parameter selection determines the classification performance of the support vector machine, so the choice of parameters is essential. Previous parameter selection of support vector machine by virtue of experience or Mongolian guess, with a lot of limitations. To overcome this drawback, Chapella parameter selection method based on gradient descent support vector machine[4],Qi Liang proposed parameter selection method of support vector machine based on ant colony algorithm[5] Wang Xingling and cars accounted for Bin-based network The lattice parameter optimization method[6], these methods in the corresponding literature experiments proved the effectiveness of various methods, but there are still limitations of these algorithms, easy to fall into local minimum point, computationally intensive, low efficiency defects.

With the in-depth development of China's market economy

and health care system reform, medical industry competition between health care workers and patients is no longer a simple treatment with the therapeutic relationship the characteristics of the service industry is gradually emerging hospital service mode from the disease to the patient transfer and hospital customer data classification, through the network to establish and maintain long-term good relations of the hospital with customers, dig and effective management of customer resources, to maintain and increase the hospital's market competitiveness has very important significance. Classification (to determine which project played a catalytic role) can play a very good reference for the future integration of hospital resources and distribution of hospital customers.

In view of this, A particle swarm pattern search algorithm is proposed to optimize the parameters, the algorithm combines the advantages of particle swarm optimization algorithm and pattern search, the algorithm in efficiency and performance than particle swarm optimization and pattern search algorithm, there are a lot The improvement for the support vector machine parameter selection provides a new and effective method. Simulation results show that the optimized SVM algorithm in the hospital customer data classification process is not only efficient, but the search for the optimal parameters to achieve a higher accuracy rate.

2. Effect on SVM algorithm and kernel function

The main idea of SVM can be summarized as two points: (1) it is analyzed for linear separable case, through the use of non-linear mapping algorithm for linear inseparable transformed into low-dimensional input space linear inseparable sample above the high dimensional feature space it linearly separable, which makes high-dimensional feature space the linear algorithm linear analysis of the non-linear characteristics of the sample as possible; (2) it is based on the structural risk minimization theory in the feature space to construct optimal partition hyperplane, makes learning to get global optimization, and satisfy certain upper bound the expected risk of the entire sample space to a probability.

SVM method is by a non-linear mapping p, the sample space is mapped to a high-dimensional, even infinite dimensional feature space (Hilbert space), so the problem into a nonlinear separable in the original sample space in the feature space in linearly separable problems. Simply put, the raised dimension and linear. Raised dimension, is mapping the samples to the high-dimensional space, under normal circumstances, this will increase the complexity of the calculation, and may even cause the "*curse of dimensionality*", and thus rarely cares. As issues such as classification, regression, however, is likely not linear processing in low-dimensional sample space sample set in the high dimensional feature space by a linear hyperplane linear division (or regression). The raised dimension will bring computing complexity, the SVM method ingenious solution to this problem: the application kernel function expansion theorem, you do not need to know the explicit expression of non-linear mapping; high dimensional feature space linear learning machine, compared with the linear model, not only almost does not increase the complexity of the calculation, but also in a way to avoid the "*curse of dimensionality*". All thanks to the expansion of the kernel function and computing theory.

Sigmoid kernel function from the neural network was shown to have good global classification performance when its parameters to obtain appropriate generalization ability is very similar to the Gaussian kernel classification results are obvious. Sigmoid kernel function three parameters and select a different combination can construct different performance classifier.

c is to control the right and wrong stars sample punishment level, the value of smaller right or wrong stars samples punishment lower degree classification hyperplane more simple, when experience larger error; if the c value tends to infinity, then all constraints that must be met in this case means that all training samples must be correctly classified, this will result in separating hyperplane complex, computationally intensive, time-consuming and more. So the choice of the value c to be combined with the specific application, to meet the classification accuracy to take a small value as possible, so that the decision function is simple. And the γ change actually is implied to change the mapping function, thereby changing the spatial distribution of the sample data sub complexity also determines the linear classification achievable minimum error [8]. The γ Value is too small, all the test samples become support vector to produce "*Overfitting Phenomenon*"; if the γ value is too large, the support vector machine classification performance will drop substantially, the classification of the test samples is almost zero, All samples are judged as a class. Decreased number of support vectors, and the correct classification ability of the classifier on the test samples would be greatly enhanced if the γ value is appropriate, usually use c and γ match.

3. PSO mode optimized SVM parameters

3.1 PSO algorithm

PSO algorithm, each member of the groups called particles, each particle in a multi-dimensional search space flight and continue according to the particle's own experience, the particle neighbor's experience or the experience of the whole group and update their own speed and position, in order to find the optimal solution. The corresponding iterative formula is:

$$v_i^l = w \times v_i + c_1 \times rand_1^l \times (pBest - p_i^l) + c_2 \times rand_2^l \times (gBest - p_i^l) \quad (1)$$

$$p_i^l = p_i^d + v_i^l \quad (2)$$

Where, w is the inertia weight; relatively large w have more global search ability, relatively small w will lead to a rapid

convergence; c_1 and c_2 are learning factors; $rand_1^l$ and $rand_2^l$ are random numbers on the interval [0,1]; p is expressed support vector machine parameters c and σ is the current value of the current position of the particle; $v \in [v_{max}, v_{min}]$ is the velocity of the particles, to determine the profile direction and the size of the next generation of c and σ.

In order to improve the speed of convergence of the algorithm, the with shrinkage factor χ speed evolution equation (equation(3)):

$$v_i^l = \chi\{w \times v_i + c_1 \times rd_1^l \times (pBest - p_i^l) + c_2 \times rd_2^l \times (gBest - p_i^l)\} \quad (3)$$

3.2 Pattern Search Algorithm

Pattern search algorithm also known as Hooke-Jeeves method, the algorithm consists of two mobile process: First, the detection of moving another mobile mode. Probing movement is moved along the direction of the coordinate axis; mode mobile is put mobile connection along adjacent two probing points. Two processes can also be understood as follows: First, determine the favorable direction of the search; acceleration process, speed up the search is in a favorable direction.

PSO algorithm than the pattern search algorithm has a wide range of global search capability; search starting from the groups, with the implicit parallelism; search using the evaluation function value inspired; convergence is fast, simple parameter adjustment; scalability, and easy to combine with other algorithms, etc. However, in the local searching ability of the algorithm late Poor, feedback information is underutilized.

The advantages of pattern search pattern search method does not require the objective function to guide, is a direct method, and the iteration is relatively simple, especially the ability of local search. Pattern search method is compared to PSO optimization algorithm has a high learning accuracy, but the number of iterations, computationally intensive, objective function is easy to fall into local minimum points and defects and poor global search capability.

In this paper, the advantages of these two algorithms, an efficient and accurate select parameters - PSO pattern search algorithm is proposed.

3.3 PSO Pattern Search Algorithm

From the global view, PSO algorithm is a viable, good robustness optimization algorithm[9]. from the local view pattern search algorithm has local convergence;There can introduce pattern search in PSO algorithm PSO to get rid of the defects fall into local minimum initial value sensitive, improve the algorithm convergence speed and accuracy. The core idea of the PSO pattern search algorithm: first with the PSO algorithm to find an optimal parameter, re-use pattern search local search in the vicinity of the optimum parameters, the search results did not meet the accuracy requirements, global search in the global scope The search cycles until you find the optimal parameters.

PSO pattern search algorithm steps are as follows:The first step, the initialization population size, the particle velocity V_i and position P_i; The second step, To calculate the particle fitness function, the classification error (formula 4) is as the particles of the fitness evaluation function.

Classification error formula:

$$1 - \frac{1}{n}\sum_{i=1}^{n}\left|\frac{T_i}{T_i + F_i}\right| \quad (4)$$

In which, T_i is the correct classification of samples; F_i is the wrong number of samples.

The third step, recording minimum fitness function value according to the formula (1) and the formula (3) update the positions and velocities of all particles; The fourth step, The three step in the decree $P_g = P_g^k \in (c, \sigma)$ as the initial value of pattern search, And given the unit vector $e^j (j = 1,..., n)$. The initial step length $\sigma^0 = (\sigma_1^0, \sigma_2^0, ..., \sigma_n^0)^T > 0$, acceleration coefficient $\gamma > 0$, coefficient of contraction $\theta \in (0, 1)$ And precision $\varepsilon > 0$, Set $k = 0, y = P_g^k$, $j = 1$; The fifth step, from y set out., in order to parallel to the unit vector $e^j (j = 1,..., n)$ axial detection mobile:

(1) if $f(y + \sigma_j^k e^j) < f(y)$, order $y = y + \sigma_j^k e^j$, Otherwise turn to (2);

(2) if $f(y - \sigma_j^k e^j) < f(y)$, order $y = y - \sigma_j^k e^j$, Otherwise order $y = y$;

The sixth step, order $P_g^{k+1} = y$, if $f(P_g^{k+1}) < f(P_g^k)$, Deal with P_g^{k+1} along the direction of the acceleration $p^k = P_g^{k+1} - P_g^k$ model of mobile, order $y = P_g^{k+1} + \gamma p^k$, $\sigma^{k+1} = \sigma^k$, $k = k + 1$, turn the fifth step, or the seventh step;The seventh step, if $|\sigma^k| \le \varepsilon$, then stop the iteration, and then output results. Otherwise $P_g^{k+1} \ne P_g^k$, order $y = P_g^{k+1}$, $\sigma^{k+1} = \sigma^k$, $k = k + 1$, turn to fifth step; when $P_g^{k+1} = P_g^k$, Make $y = P_g^{k+1}$, $\sigma^{k+1} = \theta\sigma^k$, $k = k + 1$, turn fifth step; If the maximum number of iterations $|\sigma^k| > \varepsilon$, turn to the second step.

4. SVM algorithm and its customers' Classification application

4.1 hospital customers data processing

The hospital customer data with high dimension and non-linear characteristics, dimensions, and its included noise will reduce the performance of the classifier, and the characteristics of the experimental data sets, the hospital customer classification model based on support vector machine and the wavelet transform filtering noise reduction processing on the data set.

Lead to the curse of dimensionality, in order to avoid the

high-dimensional data to support vector machines bring large computing and high data dimensionality reduction using locally linear embedding method.

4.2 PSO mode optimized SVM parameters application

Sigmoid kernel function $K(x, x_i) = \tanh(\gamma(x \bullet x_i)) + coef)$ is selected as the kernel function support vector classifier, kernel function σ is width coefficient. Initialization setting corresponding parameter range is: $\sigma \in [0, 2], c \in [1, 1000]$.

In the experiment, using PSO algorithm, pattern search algorithm and PSO pattern search algorithm for optimizing parameters and establish classification model to classify detection. Three algorithms of the learning time and accuracy was shown in table1, table 2.

pattern search algorithm has the strong local search ability and convergence speed. PSO pattern search algorithm takes into account the groups history best position, enhancing the local search ability of the algorithm, in each iteration of the process, use a pattern search algorithm to optimize and upgrade, the better P_g Lead the population to better search, improve search efficiency, thus to ensure that the PSO algorithm rapid convergence under the premise of not affecting P_g strong global search ability. The use of PSO pattern search algorithm is to find optimal parameters, the optimal parameters are taken around in other parameters are accurate, under the combination of different parameters correspond to the accuracy are shown in Table 3, in which the coordinates of the value.

PSO	Pattern search algorithm	PSO pattern search algorithm
24.55	24.05	20.68

Table 1. Study time of three algorithms (min)

PSO	Pattern search algorithm	PSO pattern search algorithm
94.22	95.85	98.32

Table 2.The detection accuracy rate of three kinds of algorithm (%)

σ	*c*			
	1	100	200	300
0.00	55.1	56.2	58.8	70.1
0.02	61.0	64.2	66.3	73.0
0.04	72.2	75.0	75.0	76.8
0.06	80.0	76.0	85.1	85.1
0.08	73.1	87.8	90.2	88.1
0.20	75.2	93.8	98.5	89.9
0.40	73.5	91.3	93.7	90.1
0.70	73.1	89.1	87.0	87.3
1.00	72.0	87.3	83.0	85.0
1.50	71.0	86.0	82.5	85.1
2.00	70.2	83.0	81.7	83.1

Table 3. c and σ Under different combinations of accurate rate curve

From table 1 and table 2 shows, the PSO algorithm training time has less training time than the pattern search algorithm, but the PSO algorithm learning accuracy without pattern search algorithms. While PSO pattern search algorithm on the training time and the accuracy is better than PSO algorithm and pattern search algorithm, it reason is in the parameter optimization process, the PSO algorithm at each iteration is all particle position and velocity of the update, the algorithm increases, the search speed rein in, especially when the particle size is relatively large, its search speed will be slower. While the

As can be seen from the figure, in the combinations of parameters of the vicinity of the optimal combination of parameters also reaches a high classification accuracy rate, but did not use PSO mode search algorithm to find the optimal parameters to achieve high classification accuracy. The one hand show that the selection of parameters a great influence on the performance of the classifier; the other hand indicate that the PSO mode search algorithm to find the optimal parameters is an effective method.

5. Conclusion

Hospital clients classified applications of PSO algorithm, pattern search algorithm and PSO mode search algorithm, the learning time and the accuracy of the simulation analysis and comparison. The results found that: PSO pattern search algorithm is applied to the parameter optimization of support vector machine, not only to avoid the limitations of relying on the experience of selected parameters, but also shorten the training time of support vector machine, but also improve the accuracy. Instances prove PSO mode search algorithm is an efficient, accurate parameter optimization method. In addition, the selection of the kernel function of the standard has not been standardized, further research direction for the selection of different kernel function of the PSO mode search algorithm analysis.

References

[1]Peng, Bai., Xi bin, Zhang., Bin, Zhanget al. (2008). The theory of support vector machine and its applications. *Xi'an: Xi'an Electronic and Science University press.*

[2] Naiyang, Deng., Yingjie,Tian.(2009). Support vector machine — theory, algorithm and development. Beijing: *Science Press.*

[3] Gang, Li., Yu, Wang., Sheng, Zhang, J. (2006). Support vector machine in the computer signal classification using. *Computer application*, 26, 6.

[4] Chapelle, O., Vapnik, V., J. (2002). Choosing multiple parameters for support vector machines *Machine Learning*, 46.

[5] Liang, Qi J. (2008). Ant colony algorithm based on support vector machine SVM parameter selection method. 4,1.

[6] Xing ling, Wang, J. (2005). Car occupies binSupport vector machines based on the kernel parameters *Journal of Ocean University of China*, 5, 35.

[7] Xiaoyun, Zhang., Yuncai, Liu., Gauss, J. (2003). Nuclear performance analysis of support vector machines. *Computer Engineering*, 29, 8.

[8] Wu, Kuoping., Wang, Shengde. (2009). Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. *Pattern Recognition*, 42, 5.

[9] Jia, Wang., Xu .Yuhong, J. (2011).Based on momentum particle swarm hybrid nuclear SVM parameter optimization method. *Computer application*, 2, 31.

Author biographies

YAN Hui-feng received a Bachelor and Master degree in computer science from the ChongQing University in China. He is a teacher in College of Mobile Telecommunications ChongQing University of Posts and Telecom, His main research interests is in the field of machine learning, data mining and Pattern Recognition. He has more than 10 publications to his credit in international journals and conferences.

WANG Wei-feng received the Mechanical and electronic engineering master's degree from Changchun University of Science and Technology, Changchun,Jilin,China,in 2003.His research interestes include Pattern recognition and artificial intelligence and Routing protocol for Wireless Sensor Networks.

LIU jie received a Bachelor and Master degree in Architecture from the ChongQing University in China. She is a teacher in School Of Civil Engineering & Architecture, Chongqing Jiaotong Univ. Her main research interests is in the field of Digitized intelligent building, data mining, intelligent control. She has more than 8 publications to his credit in international journals and conferences.