Twin Support Vector Machines Based on Quantum Particle Swarm Optimization

Shifei Ding¹,², Fulin Wu¹, Ru Nie¹, Junzhao Yu¹, Huajuan Huang¹

1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China
   Email: 562185156@qq.com
2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190 China
   Email: dingsf@cumt.edu.cn

Abstract—Twin Support Vector Machines (TWSVM) are developed on the basis of Proximal Support Vector Machines (PSVM) and Proximal Support Vector Machine based on the generalized eigenvalues (GEPSVM). The solving of binary classification problem is converted to the solving of two smaller quadratic programming problems by TWSVM. And then it gets two non-parallel hyperplanes. Its efficiency of dealing with the problems and performance are better than the traditional support vector machines. However, it also has some problems. Its own parameters are difficult to be appointed. In order to solve this problem, on the basis of in-depth study of TWSVM, this paper proposes an algorithm that is the Twin Support Vector Machines based on Quantum Particle Swarm Optimization (QPSO-TWSVM). By the use of the global searching ability of the Quantum Particle Swarm Optimization (QPSO), QPSO-TWSVM can search the optimal parameters in the global scope and avoid itself falling into the local optimum prematurely to find the values of the parameters which are the closest to the optimal parameters. QPSO-TWSVM avoids using the empirical values to appoint the parameters successfully. Compared with the traditional TWSVM, QPSO-TWSVM can appoint the parameters more accurately and avoid selecting the parameters blindly. Because of the better parameter selections, QPSO-TWSVM improves the classification accuracy of TWSVM.

Index Terms—QPSO, TWSVM, parameter optimization, binary classification

I. INTRODUCTION

Support Vector Machine[1,2] (SVM) is a machine learning method which is used to solve the binary classification problem. It was firstly proposed by Vapnik [3] et al. It is an algorithm which is based on the VC dimension theory and the principle of structural risk minimization in the statistical learning theory [4,5]. It has the features of optimization, nuclear and the best generalization ability. In recent years, it has attracted the attention of the majority of scholars and been used in many fields [6-11]. On the basis of SVM, The majority of researchers have proposed many improved algorithms. For example, in 1999, Least Squares Support Vector Machine Classifiers was proposed by Suykens [12] et al. In 2001, Fung and Mangasarian [13] proposed the algorithm of the Proximal Support Vector Machines (PSVM). It is used to solve the binary classification problem. PSVM sets a hyperplane in each type of sample points. These two hyperplanes are parallel and the distance between them must be maximized. The solution of the problem is the hyperplane which is the closest to the optimal parameters and cancels the inequality constraints in the traditional SVM. Thereafter, in 2006, on the basis of the study of PSVM, the algorithm of Proximal SVM based on Generalized Eigenvalues (GEPSVM) was proposed by Mangasarian [15] et al. GEPSVM cancels the constraint that the two hyperplanes must be parallel in PSVM. GEPSVM makes each type of sample points be as close as possible to its hyperplane and be as far as possible away from the other sample points. Further, the solving of the problem is converted to the solving of the smallest eigenvalue of the two generalized eigenvalue problems to obtain the global extremum [14]. Thereafter, on the basis of the extensive and in-depth study of PSVM and GEPSVM, in 2007, the algorithm of Twin Support Vector Machines (TWSVM) was proposed by Jayadeva [16] et al. TWSVM solves a hyperplane for each type of sample points and makes each type of sample points be as close as possible to its hyperplane and as far as possible away from another type of sample points’ hyperplane. The two hyperplanes in TWSVM have no constraint of the parallel condition. The binary classification problem is converted to two smaller quadratic programming problems by TWSVM.

After TWSVM was proposed, it has caused the attention of many scholars. Because TWSVM has the solid theoretical foundation and the superiority of solving problems, many scholars contribute to the study of TWSVM [17-19]. Although the time of the development of TWSVM is not long, there have been many achievements in the efforts of the majority of research workers. For example, Jing Chen [20] proposed the algorithm of WLSTWSVM (Weighted Least Squares TWSVM), Qi Zhiqian [21] proposed a new type of Robust Twin Support Vector Machine for pattern classification, in 2009, Xinxheng Zhang et al applied the TWSVM to the detection of MCs [22]. As a classifier, TWSVM makes decisions about the presence of MCs or...
not. The experiments show that the classifier of TWSVM is conducive to the real-time processing of CMs. In 2012, the twin support vector machines based on rough sets was proposed by Junzhao Yu [19] et al. This algorithm uses the rough sets [23] to deal with the original data sets and then uses the TWSVM to train and predict the newly generated data sets.

However, whether TWSVM or WLSTWSVM, Robust Twin Support Vector Machine for pattern classification and so on, they always have the same deficiency that the parameters can’t be quickly and accurately appointed. Although many people proposed many improved methods, all of these methods appoint the parameters on the basis of empirical values. They can’t quickly find the precise values of the parameters. The parameters which are appointed by the empirical values can’t be optimal. Because they are just the best of the empirical values and they have a relatively large gap with the values of the real optimal parameters. This paper proposes a algorithm to find the values of the optimal parameters quickly and we can use this algorithm to improve TWSVM. That is the Twin Support Vector Machines based on Quantum Particle Swarm Optimization (QPSO-TWSVM). By the use of the global searching ability of the Quantum Particle Swarm Optimization (QPSO), QPSO-TWSVM can search the optimal parameters in the global scope and avoid itself falling into the local optimum prematurely to find the values of parameters which are the closest to the optimal parameters. QPSO-TWSVM avoids using the empirical values to appoint the parameters successfully. Compared with the traditional TWSVM, QPSO-TWSVM can appoint the parameters more accurately and avoid selecting the parameters blindly. Because of the better parameter selections, QPSO-TWSVM improves the classification accuracy of TWSVM.

II. TWIN SUPPORT VECTOR MACHINE

In 2007, the algorithm of Twin Support Vector Machines (TWSVM) was proposed by Jayadeva[16] et al. The solving of binary classification problem is converted to the solving of two smaller quadratic programming problems by TWSVM[14]. And then it gets two non-parallel hyperplanes. It makes each type of sample points be as close as possible to its hyperplane and as far as possible away from another type of sample points’ hyperplane. We use A and B to represent the two hyperplanes. If a sample point is closer to A, it belongs to the category which A represents. If a sample point is closer to B, it belongs to the category which B represents. Shown in Figure 1, the two lines represent the two classified hyperplanes and the red dots and green dots represent the training points of Category 1 and Category - 1.

By solving two quadratic programming problems in the following, we can obtain the two hyperplanes of TWSVM:

\[
\begin{align*}
\text{(TWSVM1)} & \quad \min \frac{1}{2} (A_{w1} + e_1 b_1) ^T (A_{w1} + e_1 b_1) + C_1 e_1 ^T \xi \\
\text{s.t.} & \quad (B_{w1} + e_1 b_1) ^T (B_{w1} + e_1 b_1) + C_1 e_1 ^T \xi 
\end{align*}
\]

\[
\begin{align*}
\text{(TWSVM2)} & \quad \min \frac{1}{2} (B_{w2} + e_2 b_2) ^T (B_{w2} + e_2 b_2) + C_2 e_2 ^T \xi \\
\text{s.t.} & \quad (A_{w2} + e_2 b_2) ^T (A_{w2} + e_2 b_2) + C_2 e_2 ^T \xi 
\end{align*}
\]

In the above formula, \( C_1 \) and \( C_2 \) are two penalty parameters, \( e_1 \) is the \( m_1 \)-dimensional unit column vector, \( e_2 \) is the \( m_2 \)-dimensional unit column vector, \( \xi \) is the slack vector, \( A = [x_{11}^{(i)}, x_{21}^{(i)}, ..., x_{m1}^{(i)}]^T \), \( B = [x_{11}^{(i)}, x_{21}^{(i)}, ..., x_{m1}^{(i)}]^T \), \( x_{j}^{(i)} \) represents the \( j \) th sample in the \( i \) th class.

The test samples are attributed to which category depending on it that they are closer to which plane. It means that if \( x^T w_k + b_k = \min_{r \in \{1,2\}} x^T w_r + b_r \), \( x \) belongs to the \( r \) th class and \( r \in \{1,2\} \).

B. The Non-linear Mathematical Model of TWSVM

The TWSVM introduced above is used to solve the problem of the linear separability. In the following, it will be extended to the case of non-linear separability. We introduce a kernel function in the non-linear separability.

\[
K(x^*, C^*) w + b = 0 \quad \text{and} \quad K(x^*, C^*) w + b = 0
\]

Being similar to the case of the linear separability, in the following, we construct the solving of the problem:
min \frac{1}{2} \|K(A,C')w_i + e_i^n + c_i e_i^T\|_2^2 + c_i e_i^T \cdot z_e \\
subject{to} - (K(B,C')w_i + e_i^n + c_i e_i^T)[1] + c_i e_i^T \cdot z_e \geq e_i^n, z_e \geq 0 \\
min \frac{1}{2} \|K(A,C')w_i + e_i^n + c_i e_i^T\|_2^2 + c_i e_i^T \cdot z_e \\
subject{to} - (K(B,C')w_i + e_i^n + c_i e_i^T)[1] + c_i e_i^T \cdot z_e \geq e_i^n, z_e \geq 0 

In the above formula, $c_i^n$ is the unit column vector which has the same number of rows with the kernel function of $K(A,C')$, $c_i$ is the unit column vector which has the same number of rows with the kernel function of $K(B,C')$.

Being similar to the case of the linear separability, the test samples are attributed to which category depending on it that they are closer to which plane. It means that if $K(x',C')w_i + b_i = \min K(x',C')w_j + b_j$, $x'$ belongs to the $r$ th class and $r \in \{1,2\}$.

III. USING THE QPSO TO IMPROVE THE TWSVM

Although there are many improved algorithms of TWSVM and they have improved the performance of TWSVM, most of the parameters are the appointed relying on empirical values. Parameters appointed by that way have a certain arbitrariness and unreliability and we are unable to find the optimal parameters. This will cause that the solving of the problem can’t be the best state. Therefore, in this paper, we propose that we can use QPSO to find the optimal parameters. The convergence speed of QPSO is fast and QPSO is not easy to fall into the local optimum. This will be helpful for finding the optimal parameters.

A. The Thought of QPSO

The classic PSO[24] is a searching model based on the random orbit. In the searching process, the searching space of the particles in each iteration step is a limited area and this will be easy to fall into the local optimum[25]. It can’t cover the whole feasible solution space and the speed of the particles is always limited. As optimum[25]. It can’t cover the whole feasible solution area and this will be easy to fall into the local characteristics of gas nanosensors. PSO has still a lot of BP neural network for the fitting of temperature forecast, in 2012, Weiguo Zhao[28] combined PSO with Li[27] mixed PSO and SVM for the road icing affect the application of PSO. For example, in 2010, Jian algorithm of the general PSO[26]. But this does not relatively poor and this is the biggest drawback of the probability 1. So the global convergence of PSO is conducive to find the optimal parameters of SVM. QPSO is developed and improved on the basis of PSO. QPSO is a modified PSO, which is to find the particle with the best fitness in the particle swarm. The coordinates of this particle is the optimal parameters that we need. Then we use the optimal parameters we have found to determine the model of TWSVM.

B. The Principle of the Algorithm of QPSO-TWSVM

In 2010, Yuan Ren[31] et al applied PSO to the parameter optimization of SVM. The experiments show that the PSO is conducive to find the optimal parameters of SVM. QPSO is developed and improved on the basis of PSO. Used to optimize the parameters, QPSO also have a good effect.

The core idea of the QPSO-TWSVM is that it finds the optimal parameters of TWSVM by QPSO. In the algorithm of QPSO, the position coordinates of the particle represent the parameters of TWSVM. The work is just to find the particle with the best fitness in the particle swarm. The coordinates of this particle is the optimal parameters that we need. Then we use the optimal parameters we have found to determine the model of TWSVM.

In an $N$ -dimensional searching space, in the algorithm of QPSO, there is a group of $X = \{X_1, X_2, ..., X_N\}$ made up of $M$ particles which represent the potential solution of the problem. All of the particles have the same objective function to calculate their fitness value. In the algorithm of QPSO-TWSVM in this paper, the objective function is the function which is used to calculate the classification accuracy of TWSVM. The bigger the classification accuracy, the better it is. At time $t$, the position of the $i$ -th particle is $X_i(t) = [X_{i1}, X_{i2}, ..., X_{in}]$, $i = 1,2, ..., m$. The particle has no velocity vector. The best position of individual is expressed as $P_i(t) = [P_{i1}, P_{i2}, ..., P_{in}]$, $i = 1,2, ..., m$. The global best position of the group is $G(t) = [G_1, G_2, ..., G_n]$ and $G(t) = P_g(t)$, in which $g$ is the subscript of the particle which is at the global best position.
and \( g \in \{1, 2, \ldots, M \} \). The update equation of the particle position in the algorithm of QPSO is as follows:

\[
P_{ij}(t) = \phi_{ij}(t) \cdot P_{ij}(t) + \left[ 1 - \phi_{ij}(t) \right] \cdot G_{ij}(t) \cdot \varphi_{ij}(t) ~ \sim ~ U(0, 1) \]

\[
X_{ij}(t+1) = P_{ij}(t) \pm \alpha C_{ij}(t) \left[ \ln \left| u_{ij}(t) \right| \right] u_{ij}(t) ~ \sim ~ U(0, 1) \]

In the above formula,

\[
C(t) = \frac{1}{M} \sum_{i=1}^{M} p_{ij}(t) = \frac{1}{M} \sum_{i=1}^{M} P_{ij}(t), \quad \frac{1}{M} \sum_{i=1}^{M} P_{ij}(t) \ldots \frac{1}{M} \sum_{i=1}^{M} P_{ij}(t)\]

QPSO finds the global optimum position of the particles by iteration. The position coordinates of the particle which is at the global optimum position \( G_i(t) = [G_1, G_2, \ldots, G_N] \) is the parameters that we need. Then we bring the optimal parameters we have found into the TWSVM to determine the final model.

C. The Flow of Algorithm

The steps of the algorithm of QPSO-TWSVM are as follows:

Step 1 We set the population size of the particle swarm, the dimensions of location (the number of dimensions equals to the number of the parameters) and the maximum number of the iteration. Then we initialize the particle swarm randomly.

Step 2 We bring the initialized position coordinates of the particles into the TWSVM to classify the training data set. And then we will get the classification accuracy which is the fitness. According to this, we evaluate the particles and get the initial fitness value of each particle. Then we count the global best initial fitness value and record this fitness value.

Step 3 The iteration begins to find the optimal parameters. According to the formulas of \( P_{ij}(t) = \phi_{ij}(t) \cdot P_{ij}(t) + [1 - \phi_{ij}(t)] \cdot G_{ij}(t) \) and \( X_{ij}(t+1) = P_{ij}(t) \pm \alpha C_{ij}(t) \left[ \ln |u_{ij}(t)| \right] u_{ij}(t) \), it updates position coordinates of the particles continuously and calculates the fitness values of the particles. For each particle, if this fitness value is better than the fitness value of the current best position of this particle, it updates the best fitness value of the individual particle. In this iteration, if the fitness value of the particle of the best position is better than the fitness value of the current global best position, it updates the global best fitness value.

Step 4 Whether the number of the iteration has reached the maximum number of the iteration or not. If it reaches the maximum number of the iteration, it terminates the iteration. If it does not reach the maximum number of the iteration, the number of the iteration increases by 1, it jumps to Step 3 and it records the best fitness values of the individual particles and the global best fitness value.

Step 5 It records the finally obtained global optimal particle's position of \( g = (x_1, x_2, \ldots, x_N) \). The coordinate values of this particle are the values of the optimal parameters which we need to find. Bring the finally obtained values of the optimal parameters into TWSVM. Then we will get the model of QPSO-TWSVM.

Step 6 Stop the operation.

The algorithm flow chart of QPSO-TWSVM is the Figure 2.

Through this algorithm flow chart, we can intuitively understand the process of the algorithm of QPSO-TWSVM proposed in this paper. And the six steps of the algorithm described above are clearly expressed in Figure 2. This will help you understand the algorithm proposed in this paper.

IV. ANALYSIS OF THE EXPERIMENTAL RESULTS

In order to validate the algorithm proposed by this paper, we select seven commonly used data sets in the UCI machine learning database to do the test. The 80% of the data are used for training and the remaining 20% of the data are used for testing. In the linear test, we use four data sets. They are the bupa data set, the Australian data set, the Pima-Indian data set and the Sonar data set. In the non-linear test, we use three data sets. They are the ionosphere data set, the Sonar data set and the votes data set. Using the Matlab environment, all of these tests are done in the PC with 2G memory and 320G hard drive. In the algorithm of QPSO, the population size is 30, the location dimension of the particles is 2 in the linear test, the location dimension of the particles is 3 in the non-linear test, the inertia factor is 0.5, the own factor is 2.0, the global factor is 0.05, the maximum number of the iteration is 100.


A. The Linear Test

In the linear test, we use the bupa data set, the Australian data set, the Pima-Indian data set and the Sonar data set. Their data characteristics are shown in Table I:

<table>
<thead>
<tr>
<th>Data sets</th>
<th>The number of samples</th>
<th>The number of attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian</td>
<td>690</td>
<td>14</td>
</tr>
<tr>
<td>Sonar</td>
<td>208</td>
<td>60</td>
</tr>
<tr>
<td>Bupa</td>
<td>345</td>
<td>7</td>
</tr>
<tr>
<td>Pima-Indian</td>
<td>768</td>
<td>8</td>
</tr>
</tbody>
</table>

In the linear test, in the algorithm of QPSO-TWSVM, the convergence effect diagrams on the different data sets are shown in Figure 3-6. The abscissa represents the number of the iteration, the ordinate represents the fitness value or the classification accuracy:

In the linear test, in the algorithm of QPSO-TWSVM, it trains the 80% of the data to get the optimal parameters and uses the optimal parameters to determine the model of TWSVM. After the model of TWSVM is determined, it tests the remaining 20% of the data to get the corresponding classification accuracies. We compare these accuracies from QPSO-TWSVM with the accuracies from PSVM, GEPSVM and TWSVM. The comparative results are shown in Table II:

<table>
<thead>
<tr>
<th>Data sets</th>
<th>QPSO-TWSVM</th>
<th>TWSVM</th>
<th>GEPSVM</th>
<th>PSVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian</td>
<td>87.05</td>
<td>85.80</td>
<td>80.00</td>
<td>85.43</td>
</tr>
<tr>
<td>Sonar</td>
<td>86.04</td>
<td>77.26</td>
<td>72.62</td>
<td>74.51</td>
</tr>
<tr>
<td>Bupa</td>
<td>71.01</td>
<td>68.12</td>
<td>66.36</td>
<td>70.15</td>
</tr>
<tr>
<td>Pima-Indian</td>
<td>77.27</td>
<td>73.70</td>
<td>72.04</td>
<td>74.19</td>
</tr>
</tbody>
</table>

In the linear test, in the algorithm of QPSO-TWSVM, it trains the 80% of the data to get the optimal parameters and uses the optimal parameters to determine the model of TWSVM. After the model of TWSVM is determined, it tests the remaining 20% of the data to get the corresponding classification accuracies. We compare these accuracies from QPSO-TWSVM with the accuracies from PSVM, GEPSVM and TWSVM. The comparative results are shown in Table II:
B. The Non-linear Test

In the non-linear test, we use the ionosphere data set, the Sonar data set and the votes data set. Their data characteristics are shown in Table III:

In the non-linear test, in the algorithm of QPSO-TWSVM, the convergence effect diagrams on the different data sets are shown in Figure 7-9. The abscissa represents the number of the iteration, the ordinate represents the fitness value or the classification accuracy:

<table>
<thead>
<tr>
<th>Data sets</th>
<th>The number of samples</th>
<th>The number of attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonar</td>
<td>208</td>
<td>60</td>
</tr>
<tr>
<td>Ionosphere</td>
<td>351</td>
<td>34</td>
</tr>
<tr>
<td>Votes</td>
<td>435</td>
<td>16</td>
</tr>
</tbody>
</table>

From the linear and non-linear tests, we can see that the classification accuracy of QPSO-TWSVM proposed by this paper has increased significantly compared with the traditional classification algorithms. The reason for getting such significant results is that the algorithm proposed by this paper uses the QPSO to find the optimal parameters. Using the QPSO to determine the optimal parameters is not only fast, but also not easy to fall into the local optimum. It has the good global convergence and finds the parameters more precisely. This greatly improves the classification accuracy and efficiency of TWSVM.

V. CONCLUSION

In recent years, the majority of scholars have been concerned about the classification algorithm. Thanks to that, the classification algorithm has various improvements and innovation. The TWSVM rising in recent years has also been developed rapidly. But there are also a variety of defects in the TWSVM. For the problem of appointing the parameters difficulty in the TWSVM, this paper proposes the algorithm of QPSO-
TWSVM. Using the good global searching ability of QPSO and the feature of the fast convergence speed of QPSO, this algorithm can find the optimal parameters quickly and accurately. And then it determines the model. It avoids the blindness and randomness of appointing the parameters in the traditional TWSVM. As it can be seen from the experiments that this algorithm improves the classification accuracy of TWSVM. But this algorithm also has some flaws. The generalization ability of this algorithm is relatively poor. The reason is not clear at present. In the next research work, we will start the work from this point and continue to optimize this algorithm to improve the generalization ability of this algorithm so that we can further improve the classification efficiency and classification accuracy of TWSVM.

ACKNOWLEDGMENT
This work is supported by the National Key Basic Research Program of China (No. 2013CB329502), the National Natural Science Foundation of China (No.41074003), the Opening Foundation of the Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (IIP2010-1), and The Opening Foundation of Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications.

REFERENCES


Shifei Ding is a professor and Ph.D. supervisor at China University of Mining and Technology. His research interests include intelligent information processing, pattern recognition, machine learning, data mining, and granular computing et al. He has published 3 books, and more than 100 research papers in journals and international conferences. He received his B.Sc., M.Sc. degree in mathematics and computer science in 1987, 1998 respectively from Qufu Normal University, received Ph.D. degree in computer science from Shandong University of Science and Technology in 2004, and received post Ph.D. degree in computer science from Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, and Chinese Academy of Sciences in 2006.

In addition, prof. Ding is a senior member of China Computer Federation (CCF), and China Association for Artificial Intelligence (CAA). He is a member of professional committee of Artificial Intelligence and Pattern Recognition, CCF, professional committee of distributed intelligence and knowledge engineering, CAAI, professional committee of machine learning, CAAI, and professional committee of rough set and soft computing, CAAI. He is an associate Editor-in-Chief for International Journal of Digital Content Technology and its Applications (JDCTA), acts as an editor for Journal of Convergence Information Technology (JCIT), International Journal of Digital Content Technology and its Applications (JDCTA) et al.

Fulin Wu is currently a graduate student now studying in School of Computer Science and Technology, China University of Mining and Technology, and his supervisor is Prof. Shifei Ding. He received his B.Sc. degree in computer science from China University of Mining and Technology in 2012. His research interests include cloud computing, feature selection, pattern recognition, machine learning et al.

Junzhao Yu is currently a graduate student now studying at School of Computer Science and Technology, China University of Mining and Technology, and his supervisor is Prof. Shifei Ding. He received his B.Sc. degree in computer science from China University of Mining and Technology in 2011. His research interests include pattern recognition, machine learning, and twin support vector machines et al.

Huajuan Huang, born in 1984. Received her B.Sc. degree and M.Sc. degree in applied computer Technology from Guangxi University for Nationalities, Guangxi, China, in 2006 and 2009 respectively. Since 2011, she has been a Ph.D. degree candidate in applied computer technology from the China University of Mining and Technology in 2011. Her research interests include pattern recognition, machine learning, and twin support vector machines et al.