The construction of support vector machine classifier using the artificial bee colony algorithm for multi-classifications of ultrasonic supraspinatus images

Ming-Huwi Horng

Department of Computer Science and Information Engineering, National PingTung University, PingTung, Taiwan
E-mail: horng@mail.nptu.edu.tw

The setting of parameters in the support vector machines (SVM) is very important with regard to its accuracy and efficiency. In this paper, we employ the firefly algorithm to train all parameters of the SVM simultaneously, including the penalty parameter, smoothness parameter and Lagrangian multiplier. The proposed method is called the artificial bee colony-based SVM (ABC-SVM) classifier. The ABC-SVM classifier is not considered with the feature selection because the SVM together with feature selection is not suitable for being applied to a multi-class classification problem, especially for the one-against-all multi-class SVM. The experiments are on the multi-class diagnosis of ultrasonic supraspinatus images. The experimental results demonstrated that the use of ABC-SVM classifier in classifying the patterns has maximum accuracy compared with LIBSVM.

Keywords: Support Vector Machines (SVM); Artificial Bee Colony Algorithm; Ultrasonic Supraspinatus Images; LIBSVM; Multi-Classifications.

1. Introductions

The support vector machine achieves the tradeoff between the minimum training set error and the maximization of the margin based on the Vapnik-Chervonenkis theory and structural risk minimization principle. Thus it usually has the best generalization ability. In essential, the SVM is a convex quadratic programming method with which it is possible to find the global rather than the local optima. The setting of the parameters for the SVM classifier always plays a significant role, which are the penalty parameter $C$ and the smoothness parameter $\gamma$ of the radial-based function. The grid search [1] can check all combinations of the parameters, $C$ and $\gamma$, under exponentially growing sequences. In practical application, the grid search is usually vulnerable to the local optimum. In other words, if the initial parameters $C$ and $\gamma$ are far from that global optimum, the resulting SVM classifier will not work effectively.
Recently, many bio-inspired optimization algorithms such as genetic algorithm [2], particle swarm optimization [3] and firefly algorithm [4] have been applied to train the parameters of the SVM classifier together with a powerful features selection for classification. However, the Lagrangian multipliers $\alpha_i$ are still estimated by using a grid search similar to the LIBSVM [5]. However, those of the designed support vector machines present a challenge to be combined into a one-against-all support vector machine classifier, because each SVM always holds a different set of features in the multi-class classifications. This paper proposes a new algorithm based on artificial bee colony algorithm [6], in which all parameters including the $C$, $\gamma$, and the Lagrangian multiplier $\alpha_i$, are concurrently trained. In the experiments, the proposed ABC-SVM was evaluated by the classifications for multi-class classification problems. This paper is organized as follows. Materials and Methods

1.1. Support vector machines

The support vector machine has been usually used to address a binary pattern classification problem. The binary SVM constructs a set of hyper-plane in an infinite dimensional space, which can then be divided into two kinds of representations, such as the linear and nonlinear SVM. First, we consider a binary classification problem, the training data set $T=\{(x_1,y_1), (x_2,y_2), (x_3,y_3), \ldots, (x_l,y_l)\}$, $y_i \in \{-1,1\}$, $x_i \in \mathbb{R}^d$, where $x_i$ is the data point and the corresponding $y_i$ is its designed label. The $l$ denotes the number of elements in the training data set. The linear SVM finds the optimal separating margin by solving the following optimization task:

\[
\begin{align*}
&\text{Minimize} & & \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{l} \epsilon_i, & \epsilon_i \geq 0. \\
&\text{Subject to} & & y_i(w^Tx_i + b) \geq 1 - \epsilon_i, & i=1,2,\ldots,l
\end{align*}
\]

Where $C$ is a penalty value, $\epsilon_i$ are positive slack variables, $w$ is a normal vector, $b$ is a scalar quantity. The minimum problem can be reduced by using the Lagrangian multiplier $\alpha_i$, which can obtain its optimum according to the Karush Kuhn-Tucker condition. If $\alpha_i > 0$, then the corresponding data $x_i$ is called the support vector (SV), and therefore, the linear discriminate function can be expressed with the optimal hyper-plane parameters $w$ and $b$ in the following equation:

\[
f(x) = \text{sgn}(\sum_{i=1}^{l} \alpha_i y_i x_i^T x + b)
\]

The equation (1) can be transformed into equation (4) by its unconstrained dual form.

\[
\begin{align*}
&\text{Maximize} & & \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j x_i^T x_j \\
&\text{Subject to} & & 0 \leq \alpha_i \leq C, & i=1,2,\ldots,l
\end{align*}
\]
1.2. Training the nonlinear SVM using the artificial bee colony algorithm

The training of the nonlinear SVM is essentially a constrained optimization problem. The important step of using the artificial bee colony is to decide the objective function (i.e., fitness function) and the ranges of designed parameters. The designed fitness function of the firefly-SVM is expressed in the following equation.

\[
\gamma \quad (5)
\]

The constraints of the solution string are:

1. \(0 \leq a_i \leq C, i = 1, ..., N\), and \(\sum_{i=1}^{N} a_i y_i = 0\)
2. \(-15 \leq \log_2 C \leq 15\)
3. \(-8 \leq \log_2 \gamma_i \leq 8\)

Where \(N\) is the number of training data, in practices, the ABC algorithm starts with a set of solutions population in the feature space. The string representation \(Z_i\) of each solution is an important factor for the subsequent steps of the algorithm; the solution string \(S_i\) is simulated as the multidimensional vector comprising optimization parameters, including the penalty parameter, smoothness parameter and Lagrangian multipliers. It is evident that each employed bee during the course of the search modifies their food sources according to its fitness that computed by using Eq.(4). Furthermore, the employed bee performs the best solution by random walk to exchange it when the resulting solution with a better solution. The food sources of \(m\) initial solutions are generated with \(n+2\) dimensions denoted by \(D\) solution sets with \(m\) solutions.

\[
D=[Z_1, Z_2, Z_3, ..., Z_m]
\]

\[
Z_i = (a_{i1}, a_{i2}, a_{i3}, ..., a_{in}, \log_2 C, \log_2 \gamma)^T,
\]

where \(0 \leq a_{ik} \leq C, \sum_{k=1}^{N} a_{ik} y_i = 0, -15 \leq \log_2 C \leq 15\) and \(-8 \leq \log_2 \gamma \leq 8\). The \(a_{ik}\) is the multiplier of \(k\)-th training data in the \(i\)-th candidate solution; \(\log_2 C\) and \(\log_2 \gamma\) are the penalty parameter and smooth parameter of the SVM, respectively, constructed by the solution string \(Z_i\). The corresponding fitness of all solution \(Z_i\) are evaluated.

The details of the proposed ABC-SVM algorithm are thus described as follows:

**Step 1. Generate the initial population of solutions**

Generate the \(SN\) solutions \(Z_i (i=1, 2, ..., SN)\) and then evaluated their fitness based on the equation (4). Set the cycle and the trail number of each solution \(Z_i\), \(trail_i\) are equal to 0.
Step 2. Place the employed bees on their food sources.

In step 2, each employed bee protects a certain solution and modifies it by using Eq. (6) and computes the fitness value of the new solution.

$$z_{ij} = z_{ij} + \phi_{ij} \cdot (z_{ij} - z_{kj})$$  \hspace{1cm} (6)

Where \( k \in \{1, 2, \ldots, SN\} \) and \( k \neq i \) and \( j \in \{1, 2, \ldots, n + 2\} \) are randomly selected indexes. \( \phi_{ij} \) is a random number between [-1, 1]. If the fitness of the new one is better than that of the previous one, the employed bee will memorize their position and give up the old one; otherwise, the employed bee keeps the old solutions.

Step 3. Send the onlooker bees to the food sources

In step 3, we calculate the probabilities vale \( p_i \) of the solution \( z_i \) using the Eq. (7).

$$p_i = \frac{\text{fit}(z_i)}{\sum_{i=1}^{SN} \text{fit}(z_{ij})}$$  \hspace{1cm} (7)

In practical, any food source \( i \) \( (i=1, 2, \ldots, SN) \) sequentially generates a random number between [0, 1] and if this number is less than \( p_i \), a onlooker bee are sent to food source \( z_i \) and produces a new solution. If the fitness of the new solution is higher than the old one, the onlooker memorizes the new solution and shares this information with other onlooker bees. Otherwise, the new solution will be discarded. The process is repeated until all onlookers are distributed onto the food sources and produces the corresponding new solution. If the solution \( Z_i \) does not update, the corresponding \( trail_i \) increase by 1.

Step 4. Send the scouts to discover new food sources

If the position of food source does not improve through the predetermined number of \( \text{limit} \) (i.e., \( trail_i > \text{limit} \)), then the food resource \( Z_i \) is abandoned and then the employed bee becomes a scout. Assume that the abandoned source is \( z_i \), then the scout discovers a new food source to be replaced with \( z_i \). This operation can be defined as in Eq. (8).

$$z_{ij} = z_{ij}^{min} + \text{rand}(0, 1) \cdot (z_{ij}^{max} - z_{ij}^{min}), j = 1, 2, \ldots, n + 2$$  \hspace{1cm} (8)

Where the \( z_{ij}^{min} \) and \( z_{ij}^{max} \) are the upper bound and upper bound of the \( j \)-th component of all solutions. If the new solution is better than the abandoned solution, the scout will be changed into the employed bee and assigned the \( trail_i \) to be 0. These three solution selection of employed bees, onlooker bees and scouts are repeated until the termination criteria are satisfied.
2. Experimental Results and Discussions

In general, the injury of supraspinatus always causes shoulder pain, especially rotator cuff diseases. The ultrasonography is the most frequently used image modality to assess the damage from supraspinatus. According to Neer’s diagnosis standards [7], the impingement syndrome diseases of supraspinatus can be divided into three disease groups; namely, tendon inflammation, calcific tendonitis; and supraspinatus tear. Similar to the experiments in a past study [8], the five multi-class support vector machine algorithms were employed to classify these images.

<table>
<thead>
<tr>
<th>Method used</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>False negative rate (%)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1- ABC-SVM based OAA-FSVM</td>
<td>93.10</td>
<td>90.00</td>
<td>1.75</td>
<td>93.30±3.32</td>
</tr>
<tr>
<td>1. normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. inflammation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tendon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. calcific tendon</td>
<td>96.67</td>
<td>93.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. supraspinatus tear</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method 2- original OAA-FSVM trained by LIBSVM</td>
<td>83.33</td>
<td>95.56</td>
<td>3.33</td>
<td>89.10±2.49</td>
</tr>
<tr>
<td>1. normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. inflammation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tendon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. calcific tendon</td>
<td>86.67</td>
<td>95.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. supraspinatus tear</td>
<td>90.00</td>
<td>96.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These five methods were original OAA SVM (OAA-SVM), OAA fuzzy SVM (OAA-FSVM), OAA decision-tree based SVM (OAA-DTB), one-against-one voting based SVM (OAO-VB) and one-against-one directed acyclic group SVM (OAO-DAG). The experimental results of that previous work showed that the CCR of the OAA-FSVM is the best method for the classification of supraspinatus images. The original OAA-FSVM is composed of many binary support vector machines. Five powerful texture features, sum average, sum variance, mean convergence, contrast, and difference variance, were used as the features for classification. Furthermore, many measure indices, such as sensitivity, specificity and F-score, are discussed in the reference [8]. In the current experiments, we replaced this LIBSVM-based support vector machine with ABC-SVM for comparison. Table 1 shows the performance indices of
OAA-FSVM with different trained support vector machines based on the 5-fold cross-validation. Referring to Table 1 we find that the false negative using the ABC-SVM is only 1.75%, which is better than the one for LIBSVM. This means that the OAA-FSVM using the ABC-SVM as constructed basis has a lower risk for the patient in diagnosis. At the same time, the 93.3% accuracy of ABC-SVM based OAA-FSVM is superior to the original OAA-FSVM trained by LIBSVM.

3. Conclusion

In this paper, we explore the uses of the ABC-SVM for multi-class classification. Based on the results of the current experiments on the multi-class classification of ultrasonic supraspinatus images, the following conclusions can be emphasized. The ABC-SVM attempts to simultaneously train three kinds of parameters; penalty parameter, smoothness parameter and Lagrangian multiplier. Experimental results demonstrate that ABC-SVM is capable of dealing with the applications of pattern classification. The ABC-SVM may converge with the most optimal solution within a limited time when it associates with the feature selection because of its complexity. The experimental results of the classification of ultrasonic supraspinatus images reveal that the use of the ABC-SVM as the basic machine to construct the multi-class support vector machine can effectively improve the classification performances in the multi-class classification of ultrasonic supraspinatus images.

Acknowledgement

This work was supported by the Ministry of Science and Technology, ROC, under Grant Nos. MOST 104-2221-E-153-013 and MOST 105-2221-E-153-012.

References

3. U. Aich, S. Banerjee, Modeling of EDM responses by support vector machine regression with parameters selected by particle swarm optimization,
6. M. H. Horng, Soft tissue motion estimation using the artificial bee colony algorithm in the ultrasonic image sequence, presented in international conference on materials engineering and industrial application (MEIA 2015), 2015.