Structural Fuzzy Multi-class Support Vector Machine

Kai Li and Jie Li
School of Cyber Security and Computer, Hebei University, Baoding 071000, China
Email: likai@hbu.cn

Abstract. Support vector machine is a kind of generalized linear classifier that classifies data in a supervised learning manner, and its decision boundary is the maximum margin hyperplane for solving learning samples. It is used to solve the problem of binary classification. For multi-classification problems, a combination method of multiple binary classifiers is usually used to solve, but such methods are prone to generate inseparable regions of data. Therefore, on the basis of constructing a multi-class problem directly, using the pinball loss function, and introducing the structural information of different classes in the data and the role of different samples, a new support vector machine algorithm Pin-SFSimMSVM for solving multi-classification problems is proposed. It not only retains the advantages of avoiding the existence of inseparable regions and fast calculation speed of multiple types of data, but also is insensitive to noise and resampling data, and has greatly improved the accuracy. The effectiveness of the proposed algorithm is verified by experimenting on UCI standard data sets and comparing with some multi-classification algorithms.

1. Introduction
Support vector machine (SVM) is a machine learning method for classification and regression proposed by Vapnik et al. [1] in the 1990s. SVM is mainly based on statistical learning theory, which includes the idea of structural risk minimization solution and interval maximization. It is based on input samples to restore a quadratic programming problem to obtain an optimal classification hyperplane. In order to solve the problem of high computational complexity in SVM, Jayadeva et al. [2] proposed the Twin Support Vector Machine (TWSVM) in 2007. It determines two non-parallel hyperplanes by obtaining two smaller quadratic programming problems, so that each hyperplane is closer to one class and farther away from the other class. Since then, people have proposed a variety of algorithms and their applications based on SVM and TWSVM [3-8].

It can be seen that the above algorithms are all applicable to the binary classification problem. However, the multi-classification problem is more common in practical applications. Multi-class support vector machine algorithms are mainly divided into three categories now. The first category is a multi-classification method based on binary classification [9-10], which is a combination of multi-classification strategy and binary classification method, mainly including one-versus-one twins Support vector machine (OVO-TWSVM), one-versus-all twin support vector machine (OVA-TWSVM), directed acyclic graph support vector machine (DAG-TWSVM), etc. However, it is limited due to its easy existence of data inseparable regions. The second category consists of other ameliorated solutions based on the binary classification method [11-13]. This type of method has been further improved on the basis of the previous method, making the classification effect better and more significant, which mainly includes one-versus-one-versus-rest twin support vector machine (Twin-KSVC) and multi birth support vector machine (MBSVM). The third category is to directly construct methods to solve multi-classification problems. This kind of method directly constructs a single optimization problem, which...
can consider all samples at once. And by determining the decision functions of all classes at once, and maximizing the margin from each class to the rest of the classes, to distinguish the categories to which different samples belong. This method effectively solves the problem of inseparable regions in other methods. Weston and Watkins et al. [14] proposed this kind of solution, but the main drawback of this method is its high computational complexity. Crammer and Singer et al. [15] make the constraints more compact by introducing the Kroneck function. However, its computational load is still relatively large. Wang et al. [16] proposed SimMSVM on this basis. In order to reduce the complexity of the problem, new relaxed bound conditions are introduced. This makes the constraint of the algorithm from proportional to l instead of lxk. This progress greatly improves the calculation speed.

In this paper, the pinball loss function, fuzzy membership and structural information of samples are introduced into the SimMSVM algorithm, and a new multi-class support vector machine algorithm is proposed, that is, the structural fuzzy multi-class support vector machine based on pinball loss (Pin-SFSimMSVM). It can not only make good use of the potential structural information in the samples, but also give different weights to different samples to distinguish the importance of different samples, making the algorithm insensitive to noise and resampling data and better in classification. In addition, using different methods of obtaining these two kinds of information, experiments were conducted on standard data sets with different noises. Compared with other multi-classification methods, the performance of the proposed algorithm Pin-SFSimMSVM is verified.

2. Related Work

Support vector machine algorithm that can directly solve multi-classification problems is proposed in Ref. [16]. By maximizing the margin from each class to the remaining classes, all classes are distinguished, and the decision functions of all classes can be determined at once. This method not only effectively solves the problem of inseparable regions that cannot be solved by other methods, but also greatly improves the speed compared to similar methods. A k-class dataset \( T = \{ (x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l) \} \) with \( l \) samples, where \( x_i \in \mathbb{R}^d \) is the training sample, \( y_i \in \{1, \ldots, k\} \) is class label, and its optimization problem is as follows:

\[
\min_{w_m \in \mathbb{R}^d, \xi \in \mathbb{R}^l} \frac{1}{2} \sum_{m=1}^{k} w_m^T w_m + C \sum_{i=1}^{l} \xi_i \\
\text{s.t.} \quad w_m^T \varphi(x_i) - \frac{1}{k-1} \sum_{m \neq i} w_m^T \varphi(x_i) \geq 1 - \xi_i, \\
\xi_i \geq 0, \quad i = 1, \ldots, l. \tag{1}
\]

where \( C \) is the penalty parameter. For k-class problem, \( w_m \) is the classification decision vector of the \( m \)th class, and the classification decision vectors of all k classes are obtained by solving (1). For the unknown sample \( x \), its decision rule is:

\[
m_t = \arg \max_m f_m(x) = \arg \max_m \sum_{i, y_i=m} \alpha_i \kappa(x_i, x). \tag{2}
\]

The detailed description about SimMSVM is found in reference [16].

3. Structural Fuzzy Multi-Class Support Vector Machine

In this section, pinball loss function, fuzzy membership and sample structural information are introduced into SimMSVM, and a new multi-class support vector machine algorithm Pin-SFSimMSVM is proposed. This method not only retains the advantages of solving the classification problem of all samples at once in SimMSVM, which can avoid data inseparable regions and have low computational complexity, but also replaces the traditional hinge loss with Pinball loss, and adds weight and structural information to the sample. This improves the algorithm’s noise insensitivity and classification accuracy.
Given a $k$-class training data set $T = \{(x_i, y_i, S_i) | i = 1, 2, \ldots, l\}$, which contains $l$ samples, $x_i \in \mathbb{R}^n$ is the training sample data, $y_i \in \{1, 2, \ldots, k\}$ are sample labels, then the optimization problem of Pin-SFSimMSVM algorithm is:

$$
\begin{align*}
\min_{w_m, \xi} & \quad \frac{1}{2} \sum_{m=1}^{l} w_m^T w_m + C_1 \sum_{i=1}^{l} S_i \xi_i + \frac{1}{2} C_2 \sum_{m=1}^{l} w_m^T \Sigma_m w_m \\
\text{s.t.} & \quad w_m^T \varphi(x_i) - \frac{1}{k-1} \sum_{j \neq i} w_m^T \varphi(x_j) \geq 1 - \xi_i, \\
& \quad w_m^T \varphi(x_i) - \frac{1}{k-1} \sum_{j \neq i} w_m^T \varphi(x_j) \leq 1 + \frac{\xi_i}{\tau}, \quad i = 1, \ldots, l.
\end{align*}
$$

(3)

where $w_m$ is the classification decision vector of the $m$th sample, $C_1, C_2$ are parameters and $C_1 > 0, C_2 > 0$, $\xi_i$ is the error variable, $S_i$ is the fuzzy membership, $\Sigma_m$ is the structural information of the $m$th sample, and $\tau \in [0, 1]$ is the pinball loss parameter. To obtain the solution of (3), we construct the corresponding Lagrange function:

$$
L(W, \xi, \alpha, \beta) = \frac{1}{2} \sum_{m=1}^{l} w_m^T w_m + C_1 \sum_{i=1}^{l} S_i \xi_i + \frac{1}{2} C_2 \sum_{m=1}^{l} w_m^T \Sigma_m w_m - \sum_{m=1}^{l} \alpha_i \left( w_m^T \varphi(x_i) - \frac{1}{k-1} \sum_{j \neq i} w_m^T \varphi(x_j) - 1 + \xi_i \right) \\
+ \sum_{i=1}^{l} \beta_i \left( w_m^T \varphi(x_i) - \frac{1}{k-1} \sum_{j \neq i} w_m^T \varphi(x_j) - 1 - \frac{\xi_i}{\tau} \right)
$$

(4)

where $\alpha \geq 0$ and $\beta \geq 0$ are Lagrange multipliers.

Using the Karush-Kuhn-Tucker (K.K.T.) optimality conditions, we obtain:

$$
\frac{\partial L}{\partial w_m} = w_m \left( I + C_2 \Sigma_m \right) - \sum_{i \neq j=1}^{l} (\alpha_i - \beta_j) \varphi(x_i) + \frac{1}{k-1} \sum_{i \neq j=1}^{l} (\alpha_i - \beta_j) \varphi(x_j) = 0
$$

(5)

$$
\begin{align*}
 w_m &= \left( I + C_2 \Sigma_m \right)^{-1} \left( \sum_{i \neq j=1}^{l} (\alpha_i - \beta_j) \varphi(x_i) - \frac{1}{k-1} \sum_{i \neq j=1}^{l} (\alpha_i - \beta_j) \varphi(x_j) \right) \\
\frac{\partial L}{\partial \xi_i} &= C_i S - \alpha - \frac{\beta}{\tau} = 0 \\
C_i S &= \alpha + \frac{\beta}{\tau} \\
\alpha &\geq 0, \quad \beta \geq 0
\end{align*}
$$

(6)-(9)

By using the equations (7) and (9), we obtain:

$$
-\tau C_i \leq (\alpha - \beta) \leq C_i S
$$

(10)

Using the equation (4) and the above K.K.T. conditions, we can obtain the dual of (3) as follows:

$$
\min_{\gamma} \quad \frac{1}{2} \gamma^T G \gamma - e^T \gamma \\
\text{s.t.} \quad -\tau C_i \leq \gamma \leq C_i S
$$

(11)

where $\gamma= (\alpha - \beta)$, $S$ is a vector composed of membership degrees, and $G$ is a $l \times l$-dimensional matrix.
\begin{equation}
G_{i,j} = \begin{cases}
\frac{k}{k-1} (I + C_2 \Sigma_{i,j})^{-1} K_{i,j}, & y_j = y_j, \\
\frac{3k-4}{k-1} (I + C_2 \Sigma_{i,j})^{-1} K_{i,j}, & y_j \neq y_j,
\end{cases}
\end{equation}

where $K_{i,j}$ is the abbreviation of $\kappa(x_i,x_j) = \psi(x_i)^T \psi(x_j)$, $I$ is the identity matrix of appropriate dimension.

By solving its dual problem, the value of $\gamma$ is obtained, and then $w_m$, which is the classification decision vector for each class of sample, can be obtained. When classifying new samples, the samples to be classified are sequentially multiplied by the classification decision vectors of $k$ classes, the class with the largest value among all the results belongs to the class of the sample to be classified. According to the equation (6), we obtain:

\begin{equation}
f_m(x) = (I + C_2 \Sigma_m)^{-1} \left( \sum_{i,j=1}^{m} \gamma_i \kappa(x_i,x) - \frac{1}{k-1} \sum_{i,j \neq m} \gamma_i \kappa(x_i,x) \right) \tag{13}
\end{equation}

For a new sample $x$, we determine its class label by the following decision function is designed as:

\begin{equation}
m_x = \text{arg max}_{m} f_m(x) = \text{arg max}_{m} \left( (I + C_2 \Sigma_m)^{-1} \left( \sum_{i,j=1}^{m} \gamma_i \kappa(x_i,x) - \frac{1}{k-1} \sum_{i,j \neq m} \gamma_i \kappa(x_i,x) \right) \right) \tag{14}
\end{equation}

4. Experimental Results and Analysis

In this section, we evaluate the performance of the proposed algorithm. A variety of support vector machine methods for multi-class classification such as OVO-TWSVM, OVA-TWSVM, Twin-KSVC and MBSVM, and our proposed Pin-SFSimMSVM algorithm were tested in the UCI database [17] 8 standards data sets. At the same time, in order to detect the sensitivity of the proposed algorithm to noise and resampling data, we add 5% and 10% feature noise to the standard dataset, and compare their accuracy. Experimental results show that the proposed algorithm has good noise insensitivity in standard data sets. Ten cross-validation is used to find the optimal parameter value. The algorithm is tested ten times, and the average and standard deviation of the ten results are used as the final evaluation index. The grid search method is used to determine the optimal parameters. There are four parameters in the proposed algorithm, which are the penalty parameters $C_1$ and $C_2$, the gaussian kernel parameters $p$ and the pinball loss parameter $\tau$. The penalty parameters $C_1$ and $C_2$ and the gaussian kernel parameters $p$ are chosen from set $\{2^{i} | i \in [-4,-2,0,1,2,4,6,8]\}$, the pinball loss parameter $\tau$ is searched from set $\{0.1,0.2,0.5,0.8,1\}$. The Gaussian kernel function used in the experiment is $K(x,y) = \exp(-\|x-y\|^2)$.

In the experiment, eight standard data sets in the UCI database were selected to verify the algorithm, which include Iris, Zoo, Glass, Seeds, Ecoli, Balance, Soybean and CMC, respectively. In order to study the impact of different weights on the classification results of the samples, we not only compared with the algorithm without introducing weights, but also compared the test results when using three different methods to assign weight values.

Finally, the experimental results of the above five methods are compared. The results are shown in Table 1.
Table 1. Performance comparison of different algorithms using Gaussian kernel.

| Dataset | SimMSVM | Pin-SimMSVM | Pin-FSimMSVM (class-center) | Pin-FSimMSVM (fuzzy C-means) | Pin-FSimMSVM (S-curve) |
|---------|---------|-------------|-----------------------------|-----------------------------|------------------------|
| Iris    | 96.6667±2.4596 | 97.3333±1.4054 | 97.4074±2.2222 | 97.3333±4.054 | 61.3372±5.1159 |
| Zoo     | 97.1429±4.0156 | 96.2381±3.0117 | 96.761±3.8881 | 97.3333±3.0117 | 63.5659±4.2442 |
| Glass   | 92.2619±5.1159 | 93.1217±1.3084 | 93.1217±1.8613 | 93.6508±1.3084 | 64.2442±5.4517 |
| Seeds   | 85.5219±2.5252 | 85.0379±2.9123 | 85.1852±2.9123 | 85.2814±2.9123 | 86.1953±2.9123 |
| Ecoli   | 2.1105±1.9814 | 1.4308±1.8613 | 2.4089±1.9814 | 2.1909±1.8613 | 62.3256±5.4517 |
| Balance | 100±100± | 100±100± | 100±100± | 100±100± | 100±100± |
| Soybean | 52.8813±1.652 | 53.3559±1.9189 | 53.3559±2.9746 | 53.3559±2.9746 | 53.3559±2.9746 |

It can be seen that the accuracy of most data sets has been improved after using Pinball loss and fuzzy membership. As far as the three methods of obtaining sample weights are concerned, the weights obtained by clustering are relatively more stable in the algorithm, and the accuracy is better than other methods. Therefore, in the following comparison, the fuzzy membership of the algorithm is solved using the fuzzy C-means method.

In order to verify the effectiveness of our proposed algorithm Pin-SFSimMSVM, it is compared with SimMSVM, OVO-TWSVM, OVA-TWSVM, Twin-KSVC, and MBSVM, which are commonly used in multi-classification twin support vector machine algorithms, on the UCI standard data set and the experimental results are compared. At the same time, in order to test the noise insensitivity of the algorithm, 5% and 10% noise data were also added in the 8 standard data sets for comparative experiments. The noise follows a Gaussian distribution with mean 0 and variance 1. The accuracy (%) and standard deviation of the UCI standard data set under the non-linear conditions of different algorithms are shown in Table 2, where \( r \) represents the proportion of noise contained. The Pin-SFSimMSVM algorithm uses hierarchical clustering and K-means clustering methods to obtain structural information. We also used four different methods to obtain structural information and compared, the results are shown in Figure 1.

![Figure 1](image-url)
Table 2. Performance comparison of different algorithms on standard dataset and noisy dataset.

| Dataset | SimMSVM | OVO-TWSVM | OVA-TWSVM | Twin-KSVC | MBSVM | Pin-SFSimMSVM (hierarchical) | Pin-SFSimMSVM (K-means) |
|---------|---------|-----------|-----------|-----------|-------|----------------------------|-------------------------|
| Iris    | 96.6667± | 97±       | 96.3333±  | 97.0371±  | 97.3333± | 98±                        | 97.3333±                |
| r=0.05  | 93.75±   | 91.875±   | 91.875±   | 91.5625±  | 92.3611± | 94.0972±                   | 93.75±                  |
| r=0.1   | 90.4411± | 88.1818±  | 84.1754±  | 83.6364±  | 80.303±  | 90.9091±                   | 90.5303±                |
| Zoo     | 97.1429± | 95.2381±  | 96.1905±  | 96.1905±  | 98.5714± | 98.0952±                   |                         |
| r=0.05  | 93.6364± | 93.6364±  | 91.8182±  | 91.4141±  | 92.2727± | 95.4545±                   | 96.3636±                |
| r=0.1   | 91.7874± | 91.7391±  | 90.4348±  | 90.3381±  | 91.3043± | 92.1739±                   | 92.9348±                |
| Glass   | 61.3372± | 62.3256±  | 58.6563±  | 65.814±   | 65.5814± | 66.5698±                   | 64.3411±                |
| r=0.05  | 57.5309± | 56.4444±  | 47.037±   | 61.7284±  | 60.9876± | 62.7778±                   | 60.9876±                |
| r=0.1   | 55±      | 55.5556±  | 45.0521±  | 58.7963±  | 53.3854± | 59.1146±                   | 58.125±                 |
| Seeds   | 92.2619± | 92.2619±  | 93.8095±  | 93.1548±  | 93.6508± | 93.9153±                   | 94.4444±                |
| r=0.05  | 89.1111± | 89.1111±  | 89.4445±  | 90±       | 89.1358± | 89.8765±                   | 90.6667±                |
| r=0.1   | 87.0213± | 86.5957±  | 85.1064±  | 87.5745±  | 87.0213± | 87.9433±                   | 87.4668±                |
| Ecoli   | 85.5219± | 85.1515±  | 82.4916±  | 85±       | 85.1515± | 85.7954±                   | 85.3535±                |
| r=0.05  | 80.2899± | 81.5942±  | 79.8712±  | 78.5024±  | 79.7101± | 81.3043±                   | 81.5942±                |
| r=0.1   | 80.0000± | 81.1111±  | 79.8611±  | 77.5463±  | 79.7222± | 81.7901±                   | 80.4167±                |
| Balance | 85.5±    | 95.52±    | 94.16±    | 90.4±     | 95.92±   | 87.1111±                   | 88.88±                  |
| r=0.05  | 85.7576± | 92.6515±  | 87.2159±  | 86.7424±  | 92.5758± | 86.4394±                   | 87.5758±                |
| r=0.1   | 83.913±  | 90.2899±  | 84.3297±  | 82.029±   | 90.2899± | 84.8551±                   | 85.1047±                |
| Soybean | 100±     | 100±      | 100±      | 100±      | 100±     | 100±                       | 100±                    |
| r=0.05  | 100±     | 100±      | 100±      | 100±      | 100±     | 100±                       | 100±                    |
| r=0.1   | 95.4545± | 94.5455±  | 93.6364±  | 96.3636±  | 94.5455± | 97.9798±                   | 97.9798±                |
| CMC     | 52.8813± | 53.5593±  | 52.1695±  | 52.2034±  | 53.8938± | 54.8776±                   | 54.339±                 |
| r=0.05  | 49.9616± | 48.0645±  | 48.6129±  | 47.2177±  | 48.3871± | 50.7527±                   | 51.6129±                |
| r=0.1   | 49.5692± | 47.1077±  | 46.4308±  | 46.1154±  | 47±      | 50.9846±                   | 51.0154±                |
|        | 3.1826   | 2.7576    | 2.8698    | 3.0122    | 1.659    | 2.0669                     | 2.232                   |
Experimental results show that, in the 8 selected data sets, the proposed algorithm shows excellent performance in 7 data sets, especially after adding noisy data, the accuracy is higher than the other five algorithms. Although the experimental results in the Balance data set are not as accurate as other multi-classification methods, they are still greatly improved compared to the SimMSVM algorithm, indicating that the proposed algorithm is insensitive to noise and resampling data.

5. Conclusion
This paper presents a new support vector machine algorithm Pin–SFSimMSVM for multi-classification. This method uses pinball loss function on the basis of SimMSVM, and obtains the fuzzy membership of the training samples and the structural information of the samples through various methods. It not only has the advantages of no inseparable regions and fast calculation speed, but also is insensitive to noise and resampling data in the sample, and has greatly improved the accuracy. The effectiveness of the proposed algorithm is verified by comparing experiments on multi-classification algorithms such as SimMSVM, OVO-TWSVM, OVA-TWSVM, Twin-KSVC, and MBSVM on UCI standard datasets.

Acknowledgments
The authors are grateful for the supports provided by Natural Science Foundation of Hebei Province (No. F2018201060).

References
[1] Vapnik V N 2000 The Nature of Statistical Learning Theory (New York: Springer-Verlag).
[2] Jayadeva R K, Khemchandani R and Chandra S 2007 Twin support vector machine for pattern classification IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (5) 905-10.
[3] Huang H J, Wei X X and Zhou Y Q 2018 Twin support vector machines: A survey Neurocomputing 300 34-43.
[4] Ding S F, Zhang N and Zhang X K 2017 Twin support vector machine: theory, algorithm and applications Neural Computing and Applications 28 (11) 3119-3130.
[5] Tanveer M, Sharma A and Suganthan P N 2019 General twin support vector machine with pinball loss function Information Sciences 494 311-327.
[6] Gupta D, Richhariya B and Borah P 2019 A fuzzy twin support vector machine based on information entropy for class imbalance learning Neural. Comput. Appl. 24 7153-64.
[7] Xue H, Chen S and Yang Q 2008 Structural support vector machine International Symposium on Neural Networks DBLP.
[8] Chen D D, Tian Y J and Liu X H 2016 Structural nonparallel support vector machine for pattern recognition Pattern Recognition 60 296-305.
[9] Chu M, Liu X and Gong R 2018 Multi-class classification method using twin support vector machines with multi-information for steel surface defects Chemometrics and Intelligent Laboratory Systems 176 108-118.
[10] Ding S H F, Zhang J and Zhang X K 2018 Survey on multi class twin support vector machines Journal of Software 29 (1) 89-108.
[11] Xu Y, Guo R and Wang L 2013 A Twin Multi-Class Classification Support Vector Machine Cognitive Computation 5 (4) 580-588.
[12] Yang Z X, Shao Y H and Zhang X S 2013 Multiple birth support vector machine for multi-class classification Neural Computing and Applications 22 (S1) 153-161.
[13] Shi S, Ding S, Zhang Z and Jia W 2020 Energy-based structural least squares MBSVM for classification Applied Intelligence 50 (3) 681-697.
[14] Weston J and Watkins C 1999 Multi-class support vector machine Proc. Europ. Symp. Artificial Neural Networks.
[15] Crammer K and Singer Y 2001 On the algorithmic implementation of multiclass kernel-based vector machines Journal of Machine Learning Research 2 (2) 265-292.
[16] Wang Z and Xue X 2014 Multi-Class Support Vector Machine (Switzerland: Springer International Publishing).

[17] Dua D and Graff C 2019 *UCI Machine Learning Repository* http://archive.ics.uci.edu/ml.