A regularized fuzzy support vector machine

Kai Li, Lifeng Gu

College of Computer Science and Technology, Hebei University, Baoding, China, 071000

ABSTRACT

By considering the role of different samples, fuzzy support vector machine is presented based on support vector machine. However, it ignores the structural information of each class samples. In this paper, structural information of the samples is introduced to fuzzy support vector machine and obtained a structural fuzzy support vector machine model. It is converted to dual problem with quadratic programming using Lagrange method. To solve this dual problem obtains the fuzzy support vector machine classifier. Experimental results in selected datasets demonstrate the effectiveness of the proposed method.

Keywords: support vector machines, fuzzy support vector machine, structure information, within class scatter

1. INTRODUCTION

Support vector machine SVM is a machine learning method proposed by Vapnik et al [1] and its theoretical foundation is the VC dimension and structural risk minimization principle. SVM is to seek the super-plane with maximum margin and rightly separable to two class samples. To this end, some researchers deeply study support vector machines and present a lot of different support vector machines, for example L SVM [2], υ-SVM [3], LS-SVM [4]. In order to solve the traditional support vector machine susceptibility to noises or outliers, resulting in over-fitting phenomenon in the process of training, some researchers introduced the theory of fuzzy into support vector machine and present a fuzzy support vector machine [5] to solve the influence of noise or outliers on support vector machines. However, how to determine a reasonable degree of membership functions is still important problem for fuzzy support vector machine. For this purpose, researchers put forward some methods that determine the degree of fuzzy membership [6,7]. In addition, Shilton et al [8] put forward a method for degree of fuzzy membership based on iterative technology. Yang et al [9] put forward fuzzy support vector machine based on kernel fuzzy C-means clustering. Since SVM considers only separability of samples between classes and ignore structure information within the class, researchers have proposed structural support vector machine [10]-[13]. In this paper, we introduce structural information into fuzzy support vector machine and obtain a novel fuzzy support vector machine model.

2. FUZZY SUPPORT VECTOR MACHINE

Fuzzy support vector machine is a classification method proposed by the Lin et al [5]. Its purpose is to overcome the effects of noises or outliers on support vector machine by introducing the importance of sample, i.e. each data point \( x_i \) is assigned a fuzzy membership value \( s_i \) \((0 \leq s \leq 1)\). Given that \( X = \{(x_i, y_i, s_i), (x_i, y_i, s_i), \ldots, (x_i, y_i, s_i)\} \), where \( x_i \in R^n \), \( y_i \in \{+1, -1\}, i = 1, 2, \ldots, l \), and \( s_i \) is the degree of fuzzy membership for \( x_i \) belonging to \( y_i \). Optimization problem for fuzzy support vector machine is in the following:

\[
\begin{align*}
\min_{w, b, \xi} & \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} s_i \xi_i \\
\text{s.t.} & \quad y_i (w^T x_i + b) + \xi_i \geq 1, i = 1, 2, \ldots, l, \\
& \quad \xi_i \geq 0, i = 1, 2, \ldots, l
\end{align*}
\]

where \( \xi_i \) is a variable of slack, \( \xi = (\xi_1, \xi_2, \ldots, \xi_l) \) is a vector that is comprised of variable of slack \( \xi_i \), and \( C \) is a penalized factor. The dual problem for (1) is as follows
3. INTRODUCTION OF STRUCTURAL INFORMATION INTO FUZZY SUPPORT VECTOR MACHINE

3.1 The structural fuzzy support vector machine model

In the structural fuzzy support vector machine, it not only maximizes the classification margin but also minimizes the within-class scatter. To this end, the following optimization problem is obtained by introducing within-class scatter into fuzzy vector machine:

\[
\begin{align*}
\max_{\alpha} & \quad \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j x_i^T x_j \\
\text{s.t.} & \quad \sum_{i=1}^{l} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \xi_i, i = 1, 2, \ldots, l
\end{align*}
\]  

(2)

To solve the optimization problem (4) above, we construct the following Lagrange function (5) and find the extreme point of \(L(w, b, \xi, \alpha, \beta)\):

\[
L(w, b, \xi, \alpha, \beta) = \frac{1}{2} w^T \Sigma w + C \sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i (y_i (w^T x_i + b) + \xi_i - 1) - \sum_{i=1}^{l} \beta_i \xi_i.
\]  

(5)

where \(\alpha \geq 0\) and \(\beta \geq 0\) are Lagrange multiplier vector, namely \(\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_l)\) , \(\beta = (\beta_1, \beta_2, \ldots, \beta_l)\).

The extreme point of \(L(w, b, \xi, \alpha, \beta)\) satisfies the following conditions:

\[
\frac{\partial L(w, b, \xi, \alpha, \beta)}{\partial w} = \Sigma w - \sum_{i=1}^{l} \alpha_i y_i x_i = 0,
\]  

(6)

\[
\frac{\partial L(w, b, \xi, \alpha, \beta)}{\partial b} = - \sum_{i=1}^{l} \alpha_i y_i = 0,
\]  

(7)

\[
\frac{\partial L(w, b, \xi, \alpha, \beta)}{\partial \xi_i} = \xi_i - \sum_{i=1}^{l} \alpha_i y_i - C - \beta_i = 0.
\]  

(8)

By applying (6), (7) and (8) into (5), we get the following dual problem for SFSVM:
\[
\max_{\alpha} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j x_i^T \Sigma^{-1} x_j \\
\text{s.t.} \sum_{i=1}^{l} \alpha_i y_i = 0, \\
0 \leq \alpha_i \leq C x_i, i = 1, 2, \cdots, l
\]  

It can be seen that (9) is a quadratic programming problem. Suppose that solution for (9) is \( \alpha^* = (\alpha_1^*, \alpha_2^*, \cdots, \alpha_l^*) \). Then we obtain the normal direction \( w = \Sigma^{-1} \sum_{i=1}^{l} \alpha_i^* y_i x_i \) about super-plane. So, the following decision function is obtained

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

where \( b \) is obtained by K-T-T condition.

### 3.2 Structural information about data samples

For a given data set, clustering method or statistical method can be used to obtain their structural information. In statistical method, structural information is mainly the within-class scatter. In fact, statistical method is viewed as special case of clustering method, i.e. number of clusters is equal to 1 in clustering method. In the following, we only consider this case.

Suppose that a given data set \( X \) is divided by \( X^+ \) and \( X^- \), namely \( X = X^+ \cup X^- \), where \( X^+ \) and \( X^- \) represent the set of +1 class and -1 class samples, respectively. And \( m_1 \) and \( m_2 \) are mean vectors of samples about \( X^+ \) and \( X^- \), respectively. Then within-class scatter about \( X \) is computed by

\[
S_w = \sum_{i, j \in X^+} (x_j - m_1)(x_j - m_1)^T + \sum_{i, j \in X^-} (x_j - m_2)(x_j - m_2)^T.
\]

### 3.3 Determination method of fuzzy membership for each sample

How to compute fuzzy membership of each sample is an important issue. For this, researchers present some methods wherein that of Lin et al.[5] presented is commonly used. Its idea is to compute of each sample from center of class in order to reduce impact of noise or outlier on support vector machine. In the following, the detailed steps are given for computing fuzzy membership of each sample.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Compute centers of class ( x_+ ) and ( x_- ) for ( X^+ ) and ( X^- ), respectively.</td>
</tr>
<tr>
<td>2</td>
<td>Calculate the distance of each sample from the respective center of class and their maximum distances of each class as the respective radius, namely ( r_+ = \max_{x, y \in X^+}</td>
</tr>
</tbody>
</table>
| 3    | Apply centers of class and radius to obtain the fuzzy membership of each sample, namely \( s_i = \begin{cases} 
1 - ||x_i - x_+|| / (r_+ + \delta), & \text{if } y_i = +1 \\
1 - ||x_i - x_-|| / (r_- + \delta), & \text{if } y_i = -1
\end{cases} \) |

where \( \delta > 0 \) is used to prevent from \( s_i = 0 \).

### 4. Experimental result and analysis

In order to verify the validity of the proposed method, we select 11 binary classification data sets in the experiment, which are obtained from the UCI and Stalog database. The detailed characteristics are as shown in Table 1. It is worth to note that Wine is three classification’s data set, but we convert it to binary data set in the experiment. In the following experiments, for each data set, we extract 90% and 70% of data set as the training set using random method, respectively, and the remaining 10% and 30% of data set is viewed as the testing set. The obtained experimental results are mean accuracies after running 10 times.
To demonstrate the effects of structural information on fuzzy support vector machine, we randomly select 90% of data set as training set to conduct experimental study aiming at par = 9.5. Some experimental results are shown in Figure 1, in which the ordinate represents the mean accuracy of 10 times running results and standard deviation. In addition, we also compare performance of SFSVM with that of FSVM, as shown in Figure 2.

**Figure 1** Mean accuracy and standard deviation for different data set using par=9.5
It can be seen from Figure 2 that the performance of the proposed method SFSVM is superior to that of fuzzy support vector machine FSVM in the 8 data set when \( \text{par} = 9.5 \), while in data sets Australian, Bupa and Wpbc, the performance of the proposed method SFSVM is worse than that of fuzzy support vector machine FSVM. However, it can be set a proper parameter value so that the proposed method is superior to fuzzy support vector machine FSVM on most data sets. In order to show such a situation, we conduct some experiments for \( \text{par} \in [3,10] \). Some experimental results are shown in Figure 3 and Figure 4, where Figure 3 give the good mean accuracy and standard deviation and their corresponding parameter values, Figure 4 shows the corresponding mean accuracy of SFSVM and that of FSVM for different data set. It can be seen that the performance of SFSVM is superior to that of FSVM in all data sets except to dataset Bupa which is slightly decreased.

![Figure 2](image1.png)  
Figure 2 Comparison of mean accuracy with SFSVM and FSVM using \( \text{par}=9.5 \)
In order to show the effect of different proportions of training data on classification, we experiment on selected 70% of data set as training data and experimental parameters par values were 4, 4.5, 5, 5.5 and 6, respectively. Some experimental results are shown in Figures 5 and Figure 6. It can be seen from the experimental results that the performance of the proposed method SFSVM is superior to that of fuzzy support vector machine FSVM in the 9 datasets, while in the Bupa and Wine data sets the performance of the proposed method SFSVM is inferior to that of fuzzy support vector machine FSVM. However, for Wine dataset, performance both methods remains substantially the same.
In summary, fuzzy support vector machine introduced into structural information can improve the generalization performance of fuzzy support vector machines to a certain extent. However, in some data sets, such as Bupa etc., performance of SFSVM has a decline. We think that the problem can be solved by adjusting parameter value.

5. CONCLUSION

Fuzzy support vector machine FSVM is a machine learning method based on SVM by considering role of different samples. However, it only concerns the separability between-class and does not consider the structural information of the sample set. In this paper, structural information is introduced into fuzzy support vector machine and obtained the structural fuzzy support vector machine model. Using Lagrange method, an optimization problem with a constraint condition, i.e. dual problem of primal problem, is obtained. By solving the dual problem, we obtain the structural fuzzy support vector machine’s classifier. In experiment, we select the 11 datasets to demonstrate the effectiveness of the proposed method.

Acknowledgement

This work is supported by postgraduate innovation project of Hebei University under grant X2016056.

Reference


**AUTHOR**

**Kai Li** received the B.S. and M.S. degrees in mathematics department electrical engineering department from Hebei University, Baoding, China, in 1982 and 1992, respectively. He received the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2001. He is currently a Professor in college of computer science and technology, Hebei University. His current research interests include machine learning, data mining, computational intelligence, and pattern recognition.

**Lifeng Gu** received the bachelor degree in college of information science and technology from Hebei Agricultural University, Baoding, Hebei, China, in 2013, and she is currently pursuing the M.E. degree in the college of computer science and technology, Hebei University, Baoding, Hebei, China. Her research interests include machine learning, data mining, and pattern recognition.