MUTUAL INFORMATION BASED ENSEMBLE SUPPORT VECTOR DATA DESCRIPTION

FARUK SUKRU USLU, ABDULLAH BAL

Electronics and Communication Engineering Department Yildiz Technical University Istanbul 34220, Turkey
E-mail: fuslu@tekok.edu.tr, bal@yildiz.edu.tr

Abstract- Hyperspectral imagery (HSI) has spatial and detailed spectral information. Therefore, it has been used in many different areas for anomaly and target detection or classification problems. Because it consists large amount of data, effective, accurate and fast computational methods have become critical issue in machine learning. The support vector data description (SVDD) is one of the powerful methods as a one-class classifier for classification problems in machine learning area. It is a non-parametric boundary method that tries to enclose the target objects in a minimum hypersphere as much as possible. Using kernel function is one of its advantages. The kernelization makes SVDD more efficient algorithms, when the objective data is not uniformly distributed. Apart from using kernel function, ensemble methods can be also used to improve classification performance of the SVDD. Giving proper weight to each classifier before combination is one of the important part of ensemble methods. In this paper, we have offered Mutual Information (MI) between each classifier in order to use as coefficients to weighted combiners.

Keywords- Bagging; Classification; Hyperspectral imagery; Machine learning; Mutual Information; Support vector data description.

I. INTRODUCTION

In hyperspectral imaging (HSI), target detection and classification can be performed with high accuracy by using the reflected radiation collected from the material surfaces in hundreds of narrow contiguous spectral bands which is ranging from the visible to the infrared region [1]. Because it has both spatial and spectral information, hyperspectral data set can be described as a hypercube where two dimensions represent the spatial coordinates of the image, a third one corresponds to the spectral coordinate [2]. HSI has been used widely in many different areas such as search-and-rescue operations, mine detection, biomedical imagery, agricultural monitoring, geological exploration, deep-space detection etc. Because it has detailed spectral information about material, HSI has huge amount of data. So that, fast and accurate computation method is needed for high dimension data in machine learning. Generally, two different methods are offered in literature; first one which is called density estimation methods uses probability density function of data to estimate the support region [3], and the other one which is called boundary techniques focuses on the boundary of the data without assuming its shape. The boundary techniques are more efficient than density estimation methods especially when the data is as big as in HSI [4]. Support Vector Data Description (SVDD) is one of the prominent one as a one-class classifier. It is a nonparametric and powerful method which constructs a minimum hypersphere enclosing the target objects as much as possible. Addition to being nonparametric, they have the other advantages which are Sparsity: need less training sample, good Generalization: model the data in any kind of distribution and use Kernel: map the data in higher dimension [5] Although SVDD does not provide a perfect boundary in the original input space because of nonlinearity of the data. Therefore, kernel function is used to make data exhibit linear patterns in order to improve the performance. Apart from using kernel function, ensemble methods can be also used to improve classification performance of the SVDD [2]. The idea of ensemble learning is to employ multiple learners and combine their predictions based on specific fusion rules [6]. In ensemble methods, the critical part is to choose appropriate combination technique to get the more accurate outcome. Voting technique is one of most preferred one. There are two different types of voting; Majority voting in which each voter has same weight and weighted voting in which each voter has different weights for each classifier [7]. If proper weights can be assigned, weighted voting mostly has better result than majority voting. Optimization methods and heuristic approaches can be used to obtain the weights [8].

In this study, we have offered the Mutual Information (MI) based weight assigning technique to define the weights in data fusion process. Bagging is used as an ensemble technique and the proposed method has been performed on the HYDICE, Washington DC Mall hyperspectral data set. It has been shown that the proposed method have much better results than the majority voting SVDD and the conventional SVDD.

II. METHODS

A. SVDD

It was offered by Tax and Duin [9] to solve one-class problems. It is a one class form of support vector machines (SVM) which is used two class classification. The minimum boundary around the target data set is tried to obtain. The boundary shape is a hypersphere which has parameters center a and
radius \( R \). This hypersphere is used to decide whether new objects are targets or non-target. Let the \( \{y_i, i=1,\ldots,N\} \) be a set of training samples. The following optimization procedure is constructed to minimize \( R \):

\[
\min_{R, a, \xi} \left\{ R^2 + C \sum_i \xi_i \right\}
\]

(1)

with the constraints;

\[
\|\phi(y^i) - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0
\]

(2)

where \( \phi \) is a mapping function, \( \xi_i \) is slack variable to relax the boundary and \( C \) is the user-defined parameter that controls the tradeoff between errors and the volume of the hypersphere. This optimization problem can be solved using Lagrange multiplier method [10]. After solving (2), a new target \( z \) is predicted to be the target if,

\[
\|\phi(z) - a\|^2 \leq R^2
\]

(3)

The mapping function \( \phi(z) \) is utilized to transform the original data to a higher dimensional feature space to have linear patterns in higher dimensions. Because feature mapping is computationally difficult in high dimension, the kernel trick is used to make it faster. The inner products can be replaced by a kernel function, \( K(y_i, y_j) = \langle \phi(y_i), \phi(y_j) \rangle \) which should satisfy Mercer’s theorem. In this study, we have used Gaussian radial basis function (RBF) kernel which is calculated as given in eq. (4),

\[
K(x_i, x_j) = \exp\left(-\frac{\|y_i - y_j\|^2}{\sigma^2}\right)
\]

(4)

where \( \sigma \) is a width parameter to control the sensitivity of Kernel.

B. Bagging

Bagging has been proposed by Breiman, which is derived from “bootstrap aggregation” and simple method for ensemble learning [11]. Bagging is used random sampling with replacement. The purpose of bagging is to produce \( K \) replicated training data sets from the original training data set. Generally each replicated training set has the same sizes with the original training set. These \( K \) replicated training data sets are given to each classifier to get \( K \) different results. After having independent results, weighted voting is used to get final result.

\[
f_{wo} = \arg \max_j \left\{ w_j C_j \right\}
\]

(5)

where \( j \in \{-1, 1\} \), \( C_j \) is the number of SVDDs whose decisions are known to the \( j \)th class and \( w_j \) is the weight function for which mutual information is assigned in this study.

C. Mutual Information

Mutual Information is a measurement of how similar the joint distribution \( p(X,Y) \) is to the products of factored marginal distribution \( p(X)p(Y) \) [12]. Let two random variable \( X \) and \( Y \) with marginal probability distribution \( p_1(x) \) and \( p_2(y) \) and joint probability distribution \( p(x,y) \), \( x \in X \), \( y \in Y \), the mutual information of two discrete random variables \( X \) and \( Y \) can be defined as[13],

\[
I(X;Y) = \sum_y \sum_x p(x,y) \log \left( \frac{p(x,y)}{p_1(x) * p_2(y)} \right)
\]

(5)

In continuous situations, the double summation is replaced with double integral,

\[
I(X;Y) = \int_{Y \times X} p(x,y) \log \left( \frac{p(x,y)}{p_1(x) * p_2(y)} \right) dx dy
\]

(6)

The higher MI values mean that both variables are so relevant. When the MI is zero, both variables are independent.

III. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

In this experiment, we have performed proposed algorithm on Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor data set. The data is captured over the Washington, DC Mall [14]. It has seven labeled classes, 1208 lines \( X \) 307 pixels with 210 spectral bands. After removing noisy bands due to water absorption, 191 bands were used in the experiments. A false color composite image of the scene is shown in Fig. 1.

<table>
<thead>
<tr>
<th>Table I. Information of Hyperspectral Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class No.</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

Fig. 1. HYDICE DC Mall, rotated and false colored image.
In this study, we have studied all 7 classes for classification purposes. Their names and sample sizes are given in Table I. The detection accuracy is evaluated with the balanced classification rate (BCR), because data sets are mostly imbalanced in one-class classification [15]. The BCR is the average between the sensitivity and the specificity:

\[
BCR = \frac{\text{Sensitivity} + \text{Specificity}}{2}
\]  

(7)

For all the implementations of SVDD as a base learner, we have used LibSVM toolbox written by Chih-Chung Chang and Chih-Jen Lin [16].

In this algorithm, 20 different training subsets have been created randomly for bagging method and RBF kernel with sigma (\(\sigma = 0.1\)) for every training subset. The experiment has been repeated 10 times to avoid skewed results. The mean and standard deviation of BCRs calculated in each repeat for single and bagging has been computed. BCRs and performance evaluation are given in Table II for each class. Average performances for all classes have been reported to see overall classification improvement for hyperspectral images.

As seen in Table II, there is performance improvement for the seven classes. According to average performance for overall classes, it is clear that the proposed methods improve classification performance of the SVDD and make it more determined.

<table>
<thead>
<tr>
<th>Class No.</th>
<th>BCR (mean ± standard deviation in percentage)</th>
<th>Performance Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single SVDD</td>
<td>Mi Weighted Ensemble SVDD</td>
</tr>
<tr>
<td>1</td>
<td>74.36±1.95</td>
<td>75.03±1.04</td>
</tr>
<tr>
<td>2</td>
<td>71.12±15.45</td>
<td>73.89±4.16</td>
</tr>
<tr>
<td>3</td>
<td>67.37±14.29</td>
<td>73.62±5.16</td>
</tr>
<tr>
<td>4</td>
<td>64.22±10.54</td>
<td>68.52±5.88</td>
</tr>
<tr>
<td>5</td>
<td>62.53±10.89</td>
<td>63.44±4.62</td>
</tr>
<tr>
<td>6</td>
<td>71.20±15.53</td>
<td>74.24±2.29</td>
</tr>
<tr>
<td>7</td>
<td>69.39±10.18</td>
<td>70.17±6.04</td>
</tr>
</tbody>
</table>

Average Performance 2.67

CONCLUSION

In this paper, we have proposed Mutual Information (MI) as a new weight coefficient for bagging method during data fusion. We have performed our algorithm on HYDICE Washington, DC Mall data set and Support Vector Data Description has been used as a one class classifier. According to BCR values calculated in each experiment show that, the proposed method produces better results compare to conventional SVDD.

ACKNOWLEDGMENT

This research was supported by a grant from The Scientific and Technological Research Council of Turkey (TUBITAK - 112E207).

REFERENCES


Mutual Information Based Ensemble Support Vector Data Description