Active Learning with Support Vector Machines in Remotely Sensed Image Classification

Zhichao Sun¹,²,³, Zhigang Liu¹,²,³*, Suhong Liu¹,²,³, Yun Zhang¹,²,³, Bing Yang¹,²,³
¹State Key Laboratory of Remote Sensing Science (Beijing Normal University), Beijing, China
²Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Beijing Normal University
³School of Geography and Remote Sensing Science, Beijing Normal University

Abstract — Support vector machine (SVM) has been widely applied in the classification of remotely sensed image. How to reduce support vector number in SVM classifier so as to reduce classification time still an important open problem, especially in the case of mass data. To obtain fast classifier with high accuracy, an active learning schema is proposed in the SVM based image classification. Experimental results with synthetic data and multi-spectral remotely sensed images show that, compare with the SVM classifiers trained with whole training sample set in a time, the SVM classifiers obtained by active selection of training instances have much fewer support vector and can always achieve relatively higher accuracy.

Keywords—remotely sensed imagery classification; active learning; support vector machine.

I. INTRODUCTION

Supervised classification method is widely used in the remotely sensed information extraction. The consumption of the time and cost is a major classifier evaluation factor. It is particularly important when a large number of remote sensing data need to be processing. As we all know, the processing time of the classification depend heavily on the complexity of the classifier. Thus, low ratio of complexity means less time consuming.

SVM algorithm is based on the statistical learning theory and the Vapnik-Chervonenkis (VC) dimension introduced by Vladimir Vapnik [1]. SVM have been successfully applied in remote sensing images classification [2]. For support vector classification, classifiers with less support vectors (SVs) will consume less classification time.

Active learning is one of the machine learning techniques, in which representational training samples are selected to be labeled and added to the training set so as to build optimal classifier using as few training samples as possible [3]. In this paper, an active learning schema is proposed in the SVM based remotely sensed image classification.

In section 2 the support vector algorithm is briefly introduced. Then, active SVM algorithm with different sample selection strategies is described in Section 3. In Section 4 experimental results with synthetic data and multi-spectral remotely sensed images are shown and analyzed. Finally, we conclude in Section 5.

II. ALGORITHM OF SUPPORT VECTOR MACHINES

Given some training data (D)

\[
D = \{(x_i, y_i) \mid x_i \in \mathbb{R}^n, y_i \in \{-1,1\}\}_{i=1}^n
\] (1)

A hyperplane can be defined as the set of points \( x_i \) satisfying

\[
w \cdot x + b = 0
\] (2)

corresponds to the decision function

\[
f(x) = \text{sgn}(w \cdot x + b)
\] (3)

with \( w \in \mathbb{R}^n, b \in \mathbb{R} \).

The optimal hyperplane is the one that separates the instances of two classes with maximum margin [1]. This can be described by following a quadratic programming (QP) optimization problem:

Minimize:

\[
\frac{1}{2} \|w\|^2
\] (4)

Subject to:

\[y_i(w \cdot x_i + b) - 1 \geq 0, (i = 1, ..., n)\]

(5)

Usually, a soft margin classifier is used when the data are not separable. Lagrange multipliers \( \alpha_i \) can be introduced and the dual problem of the SVM is shown as follow

Maximize:

\[
\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j
\] (6)

Subject to:
\[
\sum_{i=1}^{n} \alpha_i y_i = 0 \text{ where } \alpha_i > 0
\]  

(7)

Only parts of training samples have non-zero \(\alpha_i\). These samples are named as support vectors, because only they decide the position of the classification hyperplane.

For non-linear classification, transform feature space by mapping is used to fit the maximum-margin hyperplane in the transformed feature space. In our research, we choose Radial Basis Function (RBF) kernel:

\[
k(x, x') = \exp(-\gamma \| x - x' \|^2), \gamma > 0
\]  

(8)

III. ACTIVE SVM

In this study, we are trying to reduce the number of support vectors by optimizing the training set under the premise of no accuracy lost. How to select the samples from the available training data is the main problem. Schohn and Cohn discovered that selecting samples according to the proximity to the hyperplane is effectively in document classification [4]. And the active selection of instances can significantly improve the generalization performance of a learning machine [5]. For our research, considered that samples closer to the dividing hyperplane are more effective to improve dividing hyperplane, we use an active algorithm to select samples from available training sample set (a pool of labeled samples) by considering their distance to the dividing hyperplane. To make comparison, another active algorithm is defined: add samples randomly.

For random selection, we train a small initial training set, and then select samples randomly from the available training data until no sample can be found. In active strategy, we also start with the same initial training set, then compute the distance between the other labeled samples to the hyperplane. Additional training samples are randomly selected from the labeled samples, whose distance to the dividing hyperplane are smaller then the given value. When no satisfied sample can be found, the active learning process stops. We consider that the instances far from the hyperplane have less influence to the classifier, thus we choose the instances which relatively closer to the hyperplane. The range is defined below:

\[
|D_i| < 0.3
\]  

(9)

The pseudo-code of active SVM algorithm is described below.

\begin{verbatim}
Input
T_0 : An initial training set
S: A pool of labeled samples
R: Selecting range

Begin algorithm
T \leftarrow T_0, j \leftarrow 0

\textbf{do}
\textbf{j} \leftarrow j+1
C_j \leftarrow SVMTrain(T)
M_j \leftarrow SVMValidation(C_j)
D_i \leftarrow Distance to Hyperplane (S)
S_j \leftarrow Select (|D_i| < 0.3)
S \leftarrow S-S_j
T \leftarrow T \cup S_j
\textbf{until} S_j \text{ is empty}
\textbf{return} C \leftarrow C_j, M \leftarrow \{M_j\}

End algorithm
\end{verbatim}

IV. EXPERIMENTS

Several experiments were conduct to compare the performance of the different strategies above-mentioned and evaluate the effectiveness to the active SVM algorithm. Besides the random and active strategies, we train the classifiers with all training samples were also evaluated for comparison.

We use LIBSVM tools [6] to implement training and testing data. At each experiment, 2 samples (a positive and a negative) are selected as initial training set to learn a support vector classifier.

A. Synthetic Dataset.

In synthetic dataset experiments, PRTools [7] were used to generate a dataset of two Gaussian distributed classes, which are partially overlapped (Figure 1).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Synthetic training data (left picture) and testing data (right picture) follow Gaussian distribution.}
\end{figure}
In Figure 2, test accuracy of the active learner and random learner are plotted as a function of learning times. The active learner shows a little higher accuracy and it reaches the high performance much earlier than the random one. Figure 3 shows that with the circulation of the learning, the number of support vectors increase rapidly. The active one grows faster imply higher samples selecting effectiveness then random sample selecting method. In Figure 4, the test accuracy of the active learner and random learner were plotted as a function of the number of support vectors. It is obviously shown that more support vectors are useless when the accuracy achieve a certain level.

In Table 1, the numbers of support vectors (SVs), accuracy and predict time of different methods are shown. The SVs of active learning is the number of support vectors of the classifier when the learner first achieved its peak accuracy. It is obvious that, although the accuracies are similar, the SVs differences of different classifiers are large and result in different classification speed.

### Table 1

<table>
<thead>
<tr>
<th>Method</th>
<th>SVs</th>
<th>Accuracy (%)</th>
<th>Predict time (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training with all training samples</td>
<td>390</td>
<td>83.85</td>
<td>0.0654</td>
</tr>
<tr>
<td>Random learning (max accuracy)</td>
<td>217</td>
<td>84.20</td>
<td>0.0373</td>
</tr>
<tr>
<td>Active learning (max accuracy)</td>
<td>91</td>
<td>84.20</td>
<td>0.0171</td>
</tr>
</tbody>
</table>

### B. Remotely Sensed Dataset.

Multi-spectral remote sensing data were also used in experiment. In the experiments, we study two-class classification only. Because of the space limit, results of only four groups of data sets are shown in the paper.

Figures 5-8 show the accuracy of the active and random learners on remotely sensed datasets. The active learner can achieves relatively high performance in a short period of time. The speed of the optimization of the active classifier is faster than the random one.
Figure 5. Accuracy curves for corn/wheat dataset. Top is full curve, bottom is close-up of beginning segment.

Figure 6. Accuracy curves for oats/alfalfa dataset. Top is full curve, bottom is close-up of beginning segment.

Figure 7. Accuracy curves for road/roof dataset. Top is full curve, bottom is close-up of beginning segment.

Figure 8. Accuracy curves for grass/tree dataset. Top is full curve, bottom is close-up of beginning segment.
Obviously, compare with random learner the number of support vectors grows faster for the active learner (Figure 9). That is most probably because of the effectiveness of the added samples selected by active learner.

Figure 10 shows that more support vectors are useless to the classification. When the accuracy is achieved at high level, the additional SVs are relatively uselessness for accuracy improvement the SVM classifier. These useless support vectors only severely increased the complicity of the classifier and slow down its classification speed.

Applying to remotely sensed data, the active support vector classifier perform slightly better than random learner, the two classifiers above-mentioned are both better than the one trained on all available data (Table 2). The number of support vectors is decreased noticeably. We can assert that it will save much time in the predicting processing.

<table>
<thead>
<tr>
<th>Data</th>
<th>SVs (with all samples)</th>
<th>Accuracy (with all samples) (%)</th>
<th>Max Accuracy</th>
<th>Number of SVs</th>
</tr>
</thead>
<tbody>
<tr>
<td>corn/wheat</td>
<td>141</td>
<td>99.02</td>
<td>99.06</td>
<td>24</td>
</tr>
<tr>
<td>oats/alfalfa</td>
<td>27</td>
<td>96.81</td>
<td>96.88</td>
<td>14</td>
</tr>
<tr>
<td>road/roof</td>
<td>77</td>
<td>92.11</td>
<td>92.32</td>
<td>37</td>
</tr>
<tr>
<td>grass/tree</td>
<td>32</td>
<td>97.93</td>
<td>98.14</td>
<td>10</td>
</tr>
</tbody>
</table>

V. CONCLUSION

Based on experimental results with both synthetic data and multi-spectral remotely sensed images, it can be observed that using active learners can reduce the number of support vectors of the SVM classifier dramatically. At the same time, we can achieve high-precision by using the relatively simple classifier with less support vectors, which will have higher classification speed. This will be of great use in information extraction from remotely sensed data, especially when a large number of remote sensing data need to be real-time processing or there have restricts in computing complexity. There are still some problems need to be resolved such as when should we stop adding examples to result in SVM classifier with least support vector and highest classification accuracy.
Figure 10. Number of support vectors (SVs) for active and random learners as a function of learning times on remotely sensed dataset, we only choose the close-up of beginning segment to show.

ACKNOWLEDGMENT
This work is partially supported by National Key Basic Research Program (2007CB714403), the National Natural Science Foundation of China (40701101), The National High Technology Research and Development Program of China (2006AA12Z145), National Key Project of Scientific and Technical Supporting Programs(2006BAJ09B01), Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (WKL(07)0103).

REFERENCES