Assessment of a Novel Decision and Reject Method for Multi-Class Problems in a Target Classification Framework for SAR Scenarios

Wolfgang Middelmann*, Alfons Ebert, Ulrich Thoennessen
FGAN-FOM Research Institute for Optronics and Pattern Recognition

ABSTRACT

The enhancement and improvement of classifiers for SAR signatures are a permanent challenge. The focus of this paper is the development of an integrated decision-and-reject method suitable for a kernel-machine-based target classification framework for SAR scenarios. The proposed processing chain consists of a screening process identifying ROIs with target cues, a pre-processing, and a high-performance classifier. A feasible screening method has to provide a maximum of detections namely object hypotheses while the false alarm rate is of lower interest. Therefore the quality of the following classification step significantly depends on the capability of reducing the false alarms. In complex scenarios standard approaches may classify clutter objects incorrectly as targets. To overcome this problem a novel classification scheme was developed. Class discriminating information is computed in a pre-classification step by a family of two-class kernel machines. Thus, a feature vector for an additional classification stage is provided. A comparative assessment was done using a SAR data set provided by QinetiQ. First results are given in terms of ROC curves.

Keywords: SAR, Kernel-Machine-Classification, Automatic Feature Extraction, Reject Criterion, ROC Curves

1. INTRODUCTION

Target Classification is an important component of each ATR-System. A typical processing chain consist of the identification of possible target positions, i.e. regions of interest (ROI), the provision of target descriptions (attribute vectors) and the application of a high-performance classifier. Due to their robustness and their generalization properties kernel machines e.g. Support Vector Machines (SVM), see [2,8], offer a chance for solving this classification problem very accurately and efficiently. Typically a kernel machine is designed for solving two-class problems. This means, that an appropriate decision method handling multi-class problems is needed.

![Diagram](image)

Figure 1: One-to-one classification scheme with multi-class decision heuristic

In our previous investigation [4,5] we have developed the classification scheme given in Figure 1. A one-to-one decision heuristic is used, i.e. one classifier for each pair of classes. The final classification result is deduced using a two stage approach based on class voting. At first, classes without a sufficiently high number of class votes (showing less

* Dr. Wolfgang Middelmann, FGAN-FOM, Gutleuthausstr. 1, 76275 Ettlingen, Germany; middelmann@fom.fgan.de
than 70% of possible votes) are rejected. Afterwards, based on the votes of the classifiers discriminating between the remaining classes, the three best classes were selected and directly compared with each other.

Applying this classification method to high resolution target signatures in SAR (or other) image data, a distinction between the different types of targets was achieved. A screening method only provides target hypotheses. As a consequence the classification quality depends significantly on the reject capability of the classification module reducing the false alarm rate (FAR). Therefore a controllable reject criterion was designed additionally. It is given in the high dimensional feature space implicitly defined by the kernel. In this feature space classification is realized as linear discrimination. Using the above-mentioned one-to-one-heuristic, hyperplanes \( c_{ij} \) are defining boundaries between all pairs of classes. An acceptance region is defined specifying a minimal distance \( d_{\text{min}} \) to all related hyperplanes. Two reasons are responsible rejecting a sample. Either by the parameterized reject criterion or by a class conflict including an ambiguous class voting. The reject criterion was integrated into the decision heuristic.

Applying this classification module, it is possible to reject a large number of non-trained objects. Nevertheless, some clutter objects were classified as targets. Typical examples are depicted in Figure 2 right. In parts a resemblance with the real targets represented in Figure 2 left is not discernible.

![Figure 2](image)

Figure 2  left: Trained MSTAR samples                                  right: Not rejected clutter objects

This effects are caused by the classification method which is only based on class discrimination. Thus, an additional similarity measure with respect to real targets should enable a further reduction of false alarms.

In sequel, the membership assessment was investigated by means of a one-class SVM scheme. It follows a novel method that uses class discriminating features based on results of a pre-classification stage. Comparative experiments are presented for a multi-class SAR data set.

2. **THE ONE-CLASS SVM**

Membership assessment and class discrimination are the main concepts in classification. The kernel machines used are two-class classifiers with good discrimination properties. However, a function for class membership assessment is necessary rejecting objects not belonging to the trained classes.

For each single class the training data consists of a set of target samples. By this the membership assessment task is fundamentally a 1-class classification problem. It differs from conventional classification problems in the way training the classifier. It must estimate the boundary that separates the target class from all other “unseen” samples based only on data which lies on one side of it. The problem therefore is to define this boundary in order to minimize misclassifications. This problem is also known as class-clutter-distinction.
2.1. One-class SVM - Concept

A kernel method suitable for membership assessment is the so-called 1-class SVM, see [6,7]. In our investigations, we use the RBF (Radial Basis Function) kernel. Following Byun and Lee [1] RBF kernels will be utilized in particular if there is no knowledge about the classes’ structure. The RBF kernel is defined as:

\[ K_a(x, y) = \exp\left(\frac{-||x - y||^2}{\sigma}\right). \]

The RBF kernel maps the input data onto the surface of a hypersphere in the feature space. It depends only on the distance between the support vectors and the data to be classified. Figure 3 gives a schematic view of the classifier where the hypersphere is a circle centered at the origin.

The objective of 1-class SVM is to maximize the margin between the decision hyperplane and the origin. This is equivalent to Support Vector Data Description (SVDD) [3] which finds the smallest sphere in feature space enclosing the data. The 1-class SVM training therefore computes the description of the decision hyperplane as normal vector \( w \) and bias \( b \), i.e. the distance to the origin. The dual optimization problem is

\[
\frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j K_\sigma(x_i, x_j) \rightarrow \max \alpha
\]

subject to \( \sum_{i=1}^{l} \alpha_i = 1 \), \( 0 \leq \alpha_i \leq \frac{1}{\nu} \), \( i = 1, ..., l \)

where \( x_i \) are the class samples – of the selected class. The \( \alpha_i \) are Lagrange multipliers and \( 0 < \nu < 1 \) is a parameter controlling the penalty for outliers. Thus, the decision function is

\[ f(x) = \sum_{i=1}^{l} \alpha_i K_\sigma(x_i, x) - b. \]

The bias \( b \) can be recovered by using the fact that for any \( i \) with \( 0 < \alpha_i < 1/\nu \) the corresponding \( x_i \) is a support vector and therefore \( f(x_i) = 0 \) yields. In our investigations we assume that there are only target samples in the training data set, i.e. outliers have been removed in a pre-processing step. Thus the adjustment of \( 0 < \nu < 1 \) is easy.
While using the RBF kernel, the single kernel parameter $\sigma$ is responsible for the input classes clustering fineness and gives different generalizations with respect to class-clutter-distinction. Moreover, it is possible to govern the reject by a user-defined reject threshold for the decision function adjusting the system’s selectivity.

2.2. The 1-class SVM scheme for multi-class problems

Figure 4 gives a scheme for multi-class classification based on 1-class SVMs. The 1-class SVMs are operating on the original data $o$, i.e. without feature extraction or pre-classification.

![Figure 4: 1-class SVM scheme for multi-class problems](image)

The membership assessment $c_k(o)$ of each of the $n$ classes is computed by the decision functions similar to (1)

$$c_k(o) = \sum_{j=1}^{n} w_{ij} K_{\sigma_j}(o_j, o) + w_{i0},$$

It follows the maximum decision

$$k_{win} = \arg \max_k c_k(o).$$

The class showing the highest $c_k(o)$ is accepted as final classification result. A reject will occur if the associated membership assessment $c_{win}(o)$ is below the reject threshold $TOL$, i.e. $c_{win}(o) < TOL$.

This method depends on the kernel parameter $\sigma$ and the reject threshold $TOL$ these have to be optimized. An advantage is that the informational linking of the classes occurs at the end of the process. Therefore, the classifier system can be easily upgraded with respect to new classes. The computational effort is relatively low. A drawback is that the classifier is not supported by information from class discrimination. A prerequisite of this method is a balanced scale of all membership assessments. Otherwise, objects positioned near the border lines of two classes will be presumably classified wrong. Unfortunately, there is no guarantee that the 1-class SVMs are fulfilling this demand.

3. THE NOVEL APPROACH BASED ON CLASS DISCRIMINATING FEATURES

The proposed method, sketched in Figure 5, is operating on results of decision functions provided by a pre-classifier. Therefore, normalization has to be accomplished within this preprocessing step.

The 2-class SVMs are convenient to perform this task of pre-classification. Hence, the original data $o$ is pre-classified using a scheme of 2-class SVM decision functions for discriminating all pairs of the $n$ classes resulting in a new feature vector

$$c(o) = (c_1(o),..,c_n(o))$$

$$c_j(o) = \sum_{j=1}^{n} w_{ij} K_{\sigma_j}(o_j, o) + w_{i0}, \quad i = 1,..,N$$

(2)
with \( N = n \times (n - 1) / 2 \) elements. The samples \( o_{i1} \) to \( o_{i,n(i)} \) are those of the \( i \)-th pair of classes, and \( w_{i0} \) to \( w_{i,n(i)} \) are the weights of the appropriate 2-class SVM.

Then the 1-class SVMs operate on this new feature vector \( c(o) \) with class discriminating features. They determine the membership assessment \( v_i(c(o)) \) for each of the \( n \) classes:

\[
v_i(c(o)) = \sum_{j=1}^{n(i)} u_{ij} K_n(c_{ij}, c(o)) + u_{i0}
\]

(3)

The \( c_{i1} \) to \( c_{i,n(i)} \) are those \( c(o) \) for all training samples of the \( k \)-th class, and \( u_{i0} \) to \( u_{k,n(k)} \) are the weights of the appropriate 1-class SVM.

![Figure 5: Novel classification scheme with 2-class SVMs for pre-classification](image)

The class with the highest membership assessment

\[
k_{\text{win}} = \arg \max_i v_i(c(o))
\]

has to pass the reject criterion. If \( v_{k_{\text{win}}}(c(o)) \geq TOL \) then the class membership is above the reject threshold and the sample will be assigned to this class.

An insight view of this approach is sketched in Figure 6. In this example three classes \( A \), \( B \), and \( C \) have to be discriminated. After the first stage of applying 2-class SVMs the classes are mapped into a space of class discriminating features. Especially for the membership assessment of class \( A \) two feature space coordinates are displayed in Figure 6. They are related to the decision functions (2) of the two classifiers with respect to class \( C \) vs. class \( A \) (abscissa) and class \( B \) vs. class \( A \) (ordinate). It is to be seen that the training data of class \( A \) is mapped into the positive hyperquadrant. Support vectors are projected onto the boundaries with coordinate values -1 or 1. The other class samples are lying beyond them. Thus, the training data of class \( A \) is separated from the other classes by a margin of width 2. The 1-class SVM (3) appropriate for assessing the class \( A \) membership is trained on these feature vectors of class \( A \). A possible decision boundary of class \( A \) 1-class SVM is also depicted in the figure. Its kernel parameter \( \sigma_1 \) is responsible for the clustering fineness. Therefore, the class \( A \) is split into two clusters in this example.

The following modification of the approach is possible. The separation of one class from all other classes is achieved using the results of only \((n-1)\) classifiers. The outcomes of all other ones are redundant and may even corrupt the final decision. In order to avoid this difficulty, after the pre-classification, class-dependant feature vectors \( c^{(k)} \) were generated (selected from the full-size vector \( c(o) \), see (2)). Then the 1-class SVMs are operating on these reduced feature vectors.

An advantage of both variants of the novel method is the usage of class discriminating features. The computational effort is a little greater than that of our former method using the decision heuristic. The novel method depends on the kernel parameter \( \sigma_2 \) of the 2-class SVMs and \( \sigma_1 \) the 1-class SVMs, and the reject threshold \( TOL \) of the 1-class SVMs.
4. EXPERIMENTS

A comparative assessment of the methods was accomplished using a SAR data set provided by QinetiQ. The data set consists of 4006 images subdivided into 9 classes from A to I. Each class is separated into a training set containing 335 different target aspects and into a test set containing 110 test samples. The complex-valued images of size 150x100 depict non-centered single targets. However, a fixed target position is a prerequisite for the classifier. Therefore, windows with gravity-centered targets of size 64x64 are determined in a pre-processing step. Our investigations are carried out with images alike those depicted in Figure 7. These are span images degraded to 32x32-pixel size, i.e. the polarimetric information was only used to improve the SNR.

Figure 7: Span image samples of the classes A to I (one class per column) in dB of the QinetiQ SAR dataset

An important property of a classifier is the capability to reject clutter objects not belonging to one of the trained classes. In order to test whether the proposed methods fulfill this demand, the following experiment was carried out. The four
classes A, B, D and G of the QinetiQ dataset are selected as target classes. The other five classes are defining confusion objects. After the classifier training, the test samples of all nine classes are classified.

For comparison purposes in Figure 8 the results concerning the previous heuristic method (Figure 1) are given, see [4,5]. The ROC curves are determined by varying two parameters. – The family of curves is parameterized by the reject threshold \(d_{\text{min}}\) and the curves itself are determined by varying the parameter \(\sigma\) of the RBF kernel. The results demonstrate that for distinct FAR intervals different optimal \(d_{\text{min}}\) values exist. Generally, using a larger \(d_{\text{min}}\) corresponding to a stronger reject criterion will result in a poorer classifier performance.

The diagram in Figure 9 presents the results of the 1-class SVM scheme proposed in section 2.2 with the original data as input. The ROC curves are generated by varying two parameters. – The reject threshold \(TOL\) parameterizes the family of curves and each single curve is parameterized by the RBF-kernel parameter \(\sigma_1\). The closed-world performance drops to 81.59%. This result is essentially lower than the 90.91% we got for the heuristic method. The results further indicate that it is impossible to achieve simultaneously a satisfactory classification rate and a sufficiently low FAR. For example, a FAR less than 10% is only feasible if a classification rate less than 50% is accepted. For FARs below 25%, we yield not more than CC = 70%. A classification higher than 75% is associated with a FAR essentially higher than 50%.

The so-called closed-world performance, i.e. the best classification rate without consideration of the FAR, is about 90.91%. In addition, a low FAR of 7.8% is possible with a CC of 77.05% (also 16.15% FAR vs. 86.59% CC are given).
In Figure 10 the results of our novel approach based on discriminating features are presented. The RBF-kernel parameter $\sigma_2$ of the 2-class SVMs defines the family of curves. The curves itself are obtained by simultaneously varying the parameter $\sigma_1$ of the 1-class SVMs and the reject threshold $TOL$ of the decision module.

![Figure 10: ROC curves for the novel approach based on class discriminating features, for different $\sigma_2$, CC and FAR in %](image1)

The closed-world performance is about 90.91% matching the one achieved with the previous heuristic method. However, we now found a lower FAR of 7.18 with a higher CC of 78.41%. Further, a CC greater than 80% associated with a FAR lower than 10% was achieved.

The result of our novel approach with reduced class discriminating feature vectors is depicted in Figure 11. According to Figure 10 the RBF-kernel parameter $\sigma_2$ defines the family of curves. Again, the curves itself were generated by concurrently varying the parameter $\sigma_1$ of the 1-class SVMs and the reject threshold $TOL$ of the decision module.

![Figure 11: ROC curves for the novel approach based on reduced class discriminating features, for different $\sigma_2$, CC and FAR in %](image2)

The closed-world performance of 90.23% is a little bit lower than the one of the novel method with full feature vectors (Figure 10). In comparison with the previous heuristic method, we found at a FAR of 7.44% a higher CC of 80.23%. A better result than that for full feature vectors we found at a FAR of 7.08 a higher CC of 79.77%. Furthermore a CC greater than 80% associated with a FAR lower than 8% was obtained.
5. CONCLUSION

A novel approach for multi-class classification improving the rejection of clutter objects was presented. An advantage is the usage of class discriminating features. Each class is separated from the other classes by a margin of fixed width. This boosts the classification quality of the succeeding membership assessment.

The novel method depends on the kernel parameter $\sigma_2$ of the 2-class SVMs and $\sigma_1$ the 1-class SVMs, and the reject threshold $TOL$ of the 1-class SVMs. The modification with reduced feature vectors improves the results of the previous heuristic method. Especially the classification rate at low FAR is higher.

The set of ROC points is parameterized with respect to the two parameters $\sigma_1$ and $TOL$. Surprisingly, it is approximately lying on a one-dimensional manifold. This holds for all experiments concerning the approach with reduced class discriminating feature vectors and for the experiment with lowest $\sigma_2$ concerning the approach with full size vectors. Further investigations should analyze the analytical interrelationship between these parameters.

The main advantage of the 1-class SVM scheme operating on original data is its low computational effort. However the quality depends on an equal scale of all membership assessments. A suitable adaptation to this method enabling the equalization requirement should be developed.

ACKNOWLEDGEMENTS

The authors wish to thank D. Blacknell at QinetiQ for providing the SAR data set.

REFERENCES