Designing NCTR algorithms when operating sensor conditions differ from training conditions

K.D. Copsey, R.O. Lane and A.R. Webb

QinetiQ, St Andrews Road, Malvern, UK
K.Copsey@signal.QinetiQ.com, rlane1@QinetiQ.com, A.Webb@signal.QinetiQ.com

Abstract: The basic assumption in target classifier design, based upon a training set of radar images of potential targets, is that the probability distribution from which the design sample is selected is the same as the distribution from which future objects will arise: *i.e.*, the training set is representative of targets that arise in the operating conditions. In many real-world scenarios this assumption is not valid. In this paper, we discuss some sources of variation. We then focus on a problem in radar target recognition in which the operating sensor differs from the sensor used to gather the training data and present a solution based on Bayesian image restoration.

Keywords: Bayesian, NCTR, generalisation, sensor drift, image restoration, inverse imaging, ISAR, SAR, autofocus, super-resolution, Markov chain Monte Carlo.

1. Introduction

1.1 Problem description

Typically in target classifier design we have:

1) A design or training set that is used to train a classifier.
2) A validation set, used as part of the training process, for model selection or termination of the classifier training procedure.
3) An independent test set, which is used to measure the *generalisation performance* of the classifier: the ability of the classifier to generalise to future objects.

These data sets are often gathered as part of the same trial and it is common that they are, in fact, different partitions of the same data set. For example, for Synthetic Aperture Radar (SAR) imagery of ground vehicles, the first flight pass over a configuration of targets might be used in the training set, and the second pass (over the same configuration of targets) used in the test set.

In many real world scenarios, particularly in non-cooperative target recognition (NCTR) applications, the operating conditions differ from those prevailing at the time of the test data collection. For example, sensor characteristics may drift with time or environmental conditions may change. These effects result in changes to the distributions from which target patterns are drawn. This is referred to as *population drift* [1].

The circumstances and degree of population drift vary from problem to problem. This presents a difficulty for target classifier performance evaluation since the test set may not be representative of the operating conditions and thus the generalisation performance quoted on the test set may be overly optimistic.

Designing classifiers to accommodate population drift is problem specific. In NCTR applications, there are several sources of possible discrepancy between training and operating conditions:

1) The training data may be gathered using a different sensor from that used in operation. For example, inverse synthetic aperture radar (ISAR) images of vehicles on a turntable may be gathered to train a classifier, but the operating data (*e.g.* from a weapon seeker) comprise Doppler beam sharpened (DBS) images of vehicles. These images will be at variable cross-range resolution.
2) The training data may comprise simulated radar high-resolution range (HRR) profiles, generated from point scatterer models of targets, while the operating conditions comprise measured HRR profiles [2].
3) Target configuration may differ between train and test conditions.
4) The angle of illumination in the operating conditions may differ from those represented in the training conditions.

A fuller description of these issues is provided in [3]. The specific aspect that we consider in this work is item (1) above: the difference between sensors in training and operating conditions. The practical application is one where the training data comprise ISAR images and the operating conditions comprise DBS images.

1.2 Motivation

For an NCTR system to be deployed successfully in military operations it must be robust to all the conditions that are likely to be faced. These include:

- Observation angle variations.
- Articulation (*e.g.* the rotation of a tank's turret with respect to its hull).
- Changes in target equipment fit and configuration (*e.g.* the presence of missiles or the status of a tracking radar).
- Variation between alternative models of the target.
- Camouflage, Concealment and Deception measures.
- Environmental conditions.

In the NCTR literature such conditions are often referred to as Extended Operating Conditions (EOCs) [4]. Development of classification techniques that are robust

to EOCs is, as yet, an unsolved problem. In practice, automated classification techniques that perform well across a range of EOCs are designed using large amounts of labelled training data. Unfortunately it is often difficult and expensive to collect such training data. This is a particular problem for airborne sensors, such as a DBS sensor - although wide areas can be surveyed by DBS sensors, such data can only be used for training a classifier if the targets are properly labelled, which will not be the case for data collected from routine surveillance.

In contrast to DBS data, it is relatively easy to collect large amounts of labelled ISAR data, by imaging targets on turntables. This ease of data collection makes it feasible to collect imagery from a variety of targets, in a wide range of configurations, for the full range of aspect angles. Thus, ISAR NCTR systems offer a greater potential for operation across a range of EOCs than systems developed using DBS sensor data. By developing mechanisms that enable a classifier trained using ISAR data to be applied to data from a DBS sensor, we can take advantage of the superior ability of an ISAR NCTR system to cope with a range of EOCs. This should lead to more robust classifications for the DBS sensor data.

1.3 Outline

The rest of this paper is organised as follows. Section 2 introduces the proposed Bayesian framework for generalising target classification between sensors. Section 3 briefly outlines the sensor measurement models. Algorithm specifics are detailed in Section 4, with experimental results provided in Section 5. Finally, conclusions are drawn in Section 6.

2. Bayesian approach

We denote measurements in the training (ISAR) conditions by the variable x and measurements in the operating (DBS) conditions by the variable z. We suppose that we have a training set $D = \{x_i, i = 1,\ldots,N\}$ of samples gathered under the training conditions. This training set is used to design a Bayesian classifier (*i.e.* a classifier based on probability distributions for the sensor measurements) [5], which outputs posterior class probabilities for each ISAR measurement of an object to be classified. The posterior class probabilities estimated by the classifier for an ISAR measurement x are denoted by $p(C = j \mid x, D)$, for $j = 1,\ldots,J$ where J is the number of target classes.

In accordance with our operational scenario, during operational use we only have access to a DBS measurement z (rather than the ISAR measurement x) so cannot use the Bayesian classifier directly. To proceed we require a model $p(x \mid z)$ for the relationship beween an operational sensor measurement and a training sensor measurement. Then, we can consider the expectation of the posterior class probabilities given the operational sensor measurement:

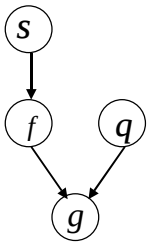
$$E[p(C = j \mid x, D) \mid z] = \int p(C = j \mid x, D)\, p(x \mid z)\, dx \qquad [1]$$

In this manner, given a model for the conditional density $p(x \mid z)$ we can use the training sensor classifier to classify an operational sensor measurement. Section 3 details the radar sensor models that ultimately enable determination of $p(x \mid z)$ for our radar NCTR application. Section 4 details the Bayesian procedures used to estimate the expectations defined in equation [1].

3. Sensor models

3.1 Measurement process

Following Luttrell [6] the overall model for the radar sensor measurement process is shown in Figure 1. A brief description is provided below, with fuller details in [6,7].



s	Scattering cross section
f	Scattered field
q	Defocus parameters
g	Image

Figure 1 : Radar sensor measurement process

Given a radar cross section s a scattered field f is generated according to a scattering model $p(f \mid s)$. The scattering model is such that each cross section element is assumed to produce a large number of wavelets that combine coherently to produce an element of the complex scattered field. Under the assumption of a sufficient number of wavelets, the central limit theorem applies, giving a complex Gaussian distribution for the scattered field conditioned on the cross section. The dependence of the cross section on the aspect angle of the target relative to the radar beam is implicit within s.

The subsequent radar imaging process can be thought of as being equivalent to applying a point spread function (PSF) to each scatterer element, and sampling at the image resolution with the addition of thermal noise. This noisy process is represented by the conditional probability distribution $p(g \mid f, q)$, where g is the radar image, and q represents the defocus parameters that alter the PSF. In our experiments $p(g \mid f, q)$ is modelled as a complex Gaussian distribution, with mean vector $T(q)f$ and noise covariance matrix N. The matrix $T(q)$ encodes the PSF. Different radar sensors (*e.g.* training and operational) can be modelled by altering the matrices $T(q)$ and N.

In the synthetic example presented in Section 5, in accordance with [6,7] the matrices $T(q)$ are defined so that the radar imaging process is equivalent to a radar platform undergoing constant cross-track acceleration in proportion to a scalar q. Specifically:

$$T(q) = \frac{1}{2c}\int_{-c}^{+c}\exp(ikx + ik^2x^2q)\,dk \approx \frac{1}{2c}\int_{-c}^{+c}\exp(ikx)(1+ik^2x^2q)\,dk = T_0 + q\,T_1 \quad [2]$$

with:

$$T_0 \equiv \frac{\sin(cx)}{cx}$$
$$T_1 \equiv i\left(cx\sin(cx) + 2\cos(cx) - \frac{2\sin(cx)}{cx}\right) \quad [3]$$

A value of $c = 1.0$ has been used for the training sensor, and $c = 0.625$ for the operational sensor. Thus, the operational sensor measurements are more distorted than the measurements from the training sensor. The defocus parameter has been set to $q = 0.1$ for both sensors. For the training sensor this value was assumed known, whereas for the operational sensor only rough prior information on the value of q was made available (reflecting the uncertainty arising from less controlled conditions during operation).

Note however, that while this choice of $T(q)$ is appropriate for a DBS sensor, it is not appropriate for an ISAR sensor (which uses target motion rather than sensor motion to build-up the sensor image). Future work will determine a PSF more appropriate for an ISAR sensor.

3.2 Measurement process

For the specified sensor measurement processes, the conditional density $p(x \mid z)$ can be expressed as:

$$p(x \mid z) = \int ds\; p(x, s \mid z) = \int ds\; p(x \mid s)p(s \mid z) \quad [4]$$

The term $p(x \mid s)$ represents the forward sensor measurement process for the training sensor, *i.e.* generation of a training sensor image from an underlying cross section s (temporarily suppressing the defocus parameters within our notation). The term $p(s \mid z)$ is the restored cross section given the operational sensor measurement. Determination of $p(s \mid z)$ corresponds to a super-resolution problem.

For a single sensor, Luttrell [6] derived a maximum likelihood technique to perform autofocus and super-resolution. The approach was applied to synthetic data generated according to the model described in Section 3.1. Given a single one-dimensional image measurement the maximum likelihood technique determined the "most likely" scattering cross section s and acceleration q. However, this does not take into account the fact that many different cross sections and defocus parameters can give rise to the same set of radar measurements, due to the interaction of several scatterers in a single pixel, defocus and the addition of noise. For this reason, a complete Bayesian approach has been proposed [7] that considers the full posterior distribution. Specifically, a Markov chain Monte Carlo (MCMC) algorithm has been developed to produce a series of samples that characterise the probability distribution of the target scattering cross section and the defocus parameters. Applied to the synthetic data, the MCMC algorithm is able to successfully determine likely values for s. However, due to the stochastic nature of the production of the scattered field, the high-probability region of the posterior distribution is typically found to be very broad. The effect of this is that large numbers of samples are needed to characterise the probability distribution, which has a detrimental effect on the computational efficiency of the algorithm. To avoid this, for our NCTR application we restore only the scattered field f. The conditional density $p(x \mid z)$ is then expressed as:

$$p(x \mid z) = \int df\; p(x, f \mid z) = \int df\; p(x \mid f)p(f \mid z) \quad [5]$$

where $p(x \mid f)$ represents the forward sensor measurement process for the training sensor, given the scattered field, and $p(f \mid z)$ is the distribution of the restored scattered field for the operational sensor measurement.

The expression for $p(x \mid z)$ in equation [5] assumes that the same scattered field will have been imaged by both the training and operational sensor. This is a simplification to the overall sensor measurement processes, in which each scattered field is generated stochastically (conditional on an underlying cross section). The distribution defined in equation [5] will therefore underestimate the uncertainty in the equivalent training sensor measurement x, because it ignores the fact that the same cross section can lead to different scattered fields. However, given enough ISAR training data, the simplified model should be sufficient.

4. Bayesian Solution

Ideally, equation [1] would be evaluated analytically using the probability distribution defined in equation [5]. However, for the majority of sensor measurement models this is not possible. Rather than making simplifying assumptions for each case, a full Bayesian approach is maintained by drawing samples from the relevant posterior distributions, and basing all inference on those samples. Specifically, the proposed Bayesian solution first draws samples from the posterior distribution $p(f, q \mid z)$. A hybrid MCMC algorithm known as Metropolis-within-Gibbs is used [8].

For a scattered field of d elements, defining:

$$f_{-i}^{(s)} = (f_1^{(s)}, \ldots, f_{i-1}^{(s)}, f_{i+1}^{(s-1)}, \ldots, f_d^{(s-1)}) \quad [6]$$

and N to be the total number of iterations, the outline of the algorithm is as follows:

1) Initialisation. Set $s = 1$ and determine initial values for $(f^{(0)}, q^{(0)})$ from the support of the joint posterior distribution.
2) Iteration s
   - For $i = 1, \ldots, d$ in turn:
     - Sample $f_i^{(s)}$ from the conditional distribution $p(f_i \mid f_{-i}^{(s)}, q^{(s-1)}, z)$ using a Metropolis-Hastings (MH) step.

- Sample $q^{(s)}$ from the conditional distribution $p(q \mid f^{(s)}, z)$ using a MH-step.

3) Increment s by 1.

- If $s > N$ then stop. Else go to step 2.

A generic MH-step for sampling from a distribution $p(h)$ works by defining a proposal distribution $q(h' \mid h)$ for updating the value of h to h'. At each iteration, a sample h' is drawn from the distribution $q(h' \mid h)$ where h is the current value of the variable. The proposed value h' is accepted with probability:

$$a(h,h') = \min\left(1, \frac{p(h')q(h \mid h')}{p(h)q(h' \mid h)}\right). \quad [7]$$

In our example Gaussian distribution random walks have been used for the proposal distributions.

By studying equation [7] we see that the MH-steps for sampling from each conditional posterior distribution $p(f_i \mid f_{-i}, q, z)$ require determination of the conditional distributions up to a normalisation constant. We write:

$$\begin{aligned} p(f_i \mid f_{-i}, q, z) &\propto p(f \mid q, z) \\ &\propto p(z \mid f, q)p(f)p(q) \end{aligned} \quad [8]$$

where $p(z \mid f, q)$ is defined in Section 3, and $p(f)$ and $p(q)$ are the prior distributions for the scattered field and defocus parameters respectively. Similarly:

$$p(q \mid f, z) \propto p(z \mid f, q)p(f)p(q) \quad [9]$$

The prior distributions $p(f)$ and $p(q)$ should be set so that the prior probability is concentrated around reasonable values. For example, if the target is in the centre of the radar beam, the prior probability for the scattered field could be set so that large amplitude values are more likely for the central elements than the outer elements. Similarly, the prior for the defocus parameters could be specified using information derived from additional sensors monitoring the platform motion.

Denoting the MCMC samples from $p(f, q \mid z)$ by $(f^{(s)}, q^{(s)})$, for $s = 1, \ldots, N$, equation [5] is rewritten as:

$$p(x \mid Z) \approx \frac{1}{N} \sum_{s=1}^{N} p(x \mid f^{(s)}) \quad [10]$$

Then, substituting into equation [1] gives:

$$E[p(C = j \mid x, D) \mid z] \approx \frac{1}{N} \int \sum_{s=1}^{N} p(C = j \mid x, D)p(x \mid f^{(s)})\, dx \quad [11]$$

Typically, this expectation is not available analytically, and is therefore evaluated using sampling techniques. Specifically, M samples are drawn from each of the N distributions $p(x \mid f^{(s)})$ (*i.e.* using the forward model for the training sensor), producing the set of samples $\{x^{(s,m)}, 1 \le s \le N, 1 \le m \le M\}$. Equation [11] can then be approximated by:

$$E[p(C = j \mid x, D) \mid z] \approx \frac{1}{NM} \sum_{s=1}^{N} \sum_{m=1}^{M} p(C = j \mid x^{(s,m)}, D) \quad [12]$$

providing the required expectations of the posterior class probabilities.

5. Results

5.1 Data

A two-class problem is used to illustrate the approach, with the classes defined using underlying cross sections of 17 elements, displayed in Figure 2. The first class corresponds to a single point scatterer, while the second class corresponds to two scatterers.

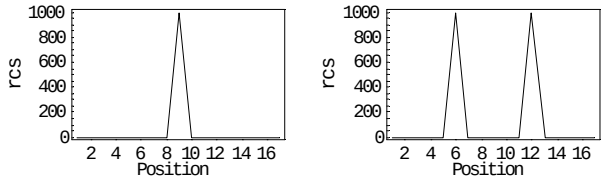


Figure 2: Underlying radar cross sections (LHS class 1, RHS class 2)

The two underlying cross sections have been used to generate scattered fields, in accordance with the complex Gaussian model introduced in Section 3.1 (and covered in more detail in [6,7]). Simulated sensor measurements have then been obtained by applying the appropriate sensor measurement model, with the resultant images being sampled at the same locations as the underlying cross section (producing one-dimensional complex images with 17 pixels). For both the training and operational sensors, independent Gaussian noise has been added to the measurements so that the signal to noise ratio (SNR) is 20 dB.

Examples of the sensor measurements for both classes (based for clarity on idealised realisations of the scattered field) are provided in Figure 3 for the training sensor and Figure 4 for the operational sensor. Note that for class 2, due to the wide PSF of both sensors relative to the separation of the two scatterers it is not immediately apparent from the images that two scatterers are present.

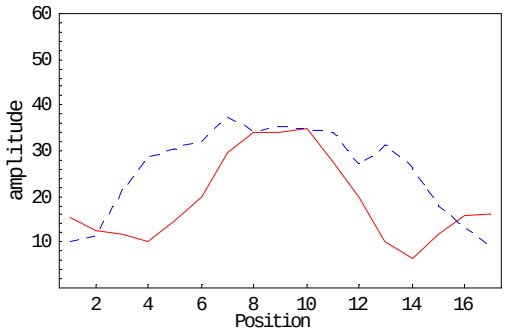


Figure 3: Training sensor measurements (solid line is an example from class 1, dashed line is from class 2)

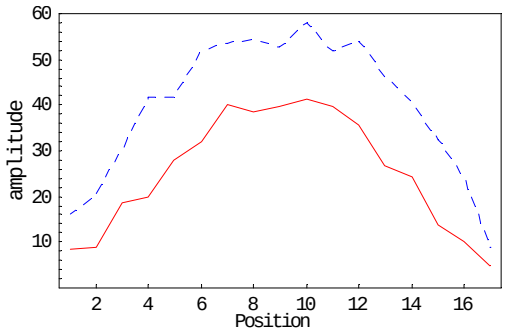


Figure 4: Operational sensor measurements (solid line is an example from class 1, dashed line is from class 2)

For each sensor the measurements from the two classes appear quite distinct (as can be seen in Figure 3 and Figure 4). Thus (for the idealised realisations of the scattered fields at least) a classifier trained using data from the training sensor should be able to classify training sensor data from the two classes. Similarly, a classifier trained using data from the operational sensor should be able to separate operating sensor data into the two classes. However, for each class there is a considerable difference in the measurements between the two sensors. Visually, the operational sensor measurement from class 1 is more similar to the training sensor measurement from class 2, than it is to the training sensor measurement from class 1. This indicates that using the training sensor classifier on the operational sensor data will lead to poor classification performance for class 1.

5.2 Experimental description

A multivariate Gaussian classifier, which outputs posterior class probabilities, has been used for the training sensor classifier [5]. A Gaussian classifier is quite a simple classifier, but is sufficient for this particular two-class problem. It is likely that more sophisticated classification techniques will be necessary for radar data of real targets. The classifier was trained using 100 training sensor measurements from each class. Pre-processing of the measurements prior to their use in the classifier consisted of taking the amplitudes of the complex measurements, and normalising each input vector so that the average amplitude was one. Note, however that for the operational sensor such pre-processing would remove much of the separability between the classes (see Figure 4).

The test data consisted of 100 operational sensor measurements from each class. The Bayesian algorithm outlined in Section 4 was applied separately to each operational sensor measurement, with each measurement being assigned to the class that maximised the estimated posterior class probability in equation [12].

Specifically, for each operational sensor measurement, 10000 samples were drawn from the posterior distribution $p(f,q \mid z)$ using the MCMC algorithm. The first 5000 of these samples were discarded as a burn-in period, and of the next 5000 only 1 in every 5 were actually stored (this sub-sampling reduces the memory storage and processing requirements). For each of the 1000 stored scattered field samples, a single sample was drawn from the training sensor measurement model $p(x \mid f)$, leading to 1000 samples from $p(x \mid z)$ for use in equation [12].

The prior distribution for q was set to be a Gaussian distribution with mean value 0.1 (the correct value) and standard deviation 0.03. The prior distribution for f was set to be a product of independent Gaussian distributions for the real and imaginary parts of each complex element. The prior means of each of the real and imaginary parts were set to zero. The prior standard deviations of the real and imaginary parts of the middle seven elements of the scattered field were all set to 10.0, with the standard deviations of the remaining parts set to 1.0. Thus, the prior distribution for the scattered field incorporated prior knowledge that large magnitude elements are unlikely, and that the middle elements are likely to be larger than the outer elements.

5.3 Experimental results

For comparison, baseline results have been obtained by applying the Gaussian classifier, trained using data from the training sensor, directly to the operational sensor data. The two sets of results are compared in Table 1.

Table 1. Correct classification rates

	Class 1	Class 2	Average
Baseline	3%	100%	51.5%
Bayesian	97%	100%	98.5%

The baseline results clearly show that ignoring the change in sensor is undesirable. In particular, the observation in Section 4.1 that the operational sensor measurements from class 1 are more similar to the class 2 training sensor measurements than those from class 1, is evidenced by 97% of the class 1 measurements being incorrectly assigned to class 2. In contrast, the proposed Bayesian procedures recover the class discriminability that is present for each specific sensor. Thus, for this example we have successfully demonstrated procedures that (given appropriate sensor measurement models) enable a classifier to be used in situations where the operating sensor differs from the sensor used to collect the training data.

6. Summary

The basic assumption in NCTR is that the training set is representative of the targets and conditions that will arise during operation. In many real-world scenarios this assumption is not valid. This paper has discussed some of the sources of variation, and then focused on a problem in radar target recognition in which the operating sensor differs from the sensor used to gather the training data. A solution based on Bayesian image restoration has been proposed, that utilises stochastic models for the radar imaging processes to convert the operating sensor measurements to have the form of the training sensor measurements. The approach is successfully illustrated for a synthetic example. Current work is extending the

models for the radar imaging processes, with a view to applying the approach to real DBS and ISAR imagery.

7. Acknowledgment

This research was sponsored by the UK MOD Corporate Research Programme.

8. References

[1] Hand, D.J. (1997). *Construction and Assessment of Classification Rules*. John Wiley, Chichester.

[2] Zwart, J.P., van der Heiden, R., Gelsema, S. and Groen, F. (2003). Fast translation invariant classification of HRR range profiles in a zero phase representation. IEE Proceedings on Radar, Sonar and Navigation, 150(6), 411-418, December 2003.

[3] Copsey, K.D. and Webb, A.R. (2004). Classifier design for population and sensor drift. Joint IAPR International Workshops SSPR 2004 and SPR 2004.

[4] Mossing, J.C. and Ross, T.D. (1998). An evaluation of SAR ATR algorithm performance sensitivity to MSTAR extended operating conditions. Algorithms for Synthetic Aperture Radar V, Proc. SPIE Vol. 3370.

[5] Webb, A.R. (2002). *Statistical Pattern Recognition*. John Wiley, Chichester, 2nd edition.

[6] Luttrell, S.P. (1990). A Bayesian derivation of an iterative autofocus/super-resolution algorithm. Inverse Problems, 6, 975-996.

[7] Lane, R.O., Copsey, K.D. and Webb, A.R. (2004). A Bayesian Approach to simultaneous autofocus and super-resolution. Algorithms for Synthetic Aperture Radar XI, Proc. SPIE Vol. 5427.

[8] Robert, C.P. and Casella G. (1999). *Monte Carlo Statistical Methods*. Springer.

9. Glossary

DBS: Doppler Beam Sharpened

EOC: Extended Operating Condition

ISAR: Inverse Synthetic Aperture Radar

MCMC: Markov chain Monte Carlo

MH: Metropolis-Hastings

NCTR: Non-cooperative Target Recognition

PSF: Point spread function

RCS: Radar Cross Section

SAR: Synthetic Aperture Radar