1. Introduction

The extent of the unexploded ordnance (UXO) contamination within the United States and abroad has motivated intensive research into improved technologies for detection and discrimination of UXO. Remediating UXO-contaminated sites in the United States could cost up to $52 billion if metal detectors and pin flags are used to locate potential UXO items (the “mag and flag” technique). This cost is distributed throughout the remediation process, but the majority of the expense is spent on digging non-ordnance items (Defense Science Board, 2003). Mag and flag can require as many as 100 clutter items (e.g., geology, shrapnel) to be excavated for each UXO item found. This 100:1 rate is called a False Alarm Rate (FAR) and is a crucial measure for comparing discrimination techniques. Advanced discrimination methods are expected to reduce false alarm rates to 10:1, cutting remediation costs to approximately $16 billion. These methods require the acquisition of digital geophysical data and subsequent signal processing.

This article provides an overview of the signal processing which has been used to discriminate between UXO and clutter. We describe classification algorithms which have been applied to this problem and demonstrate their application to a real data set. Our aim is to make these algorithms more accessible and less mysterious for the UXO community.

2. Advanced discrimination

Before we can apply advanced discrimination techniques, we must acquire digital geophysical data. Typically magnetic and/or electromagnetic (either time or frequency-domain) data are collected. Once anomalies have been identified in the observed data, we can characterize each one by finding the model parameters which best predict the observed data. For magnetic data we generally attempt to find a best-fitting dipole. Six parameters specifying location, strength and orientation are sought by solving an overdetermined inverse problem, that is, a problem which has more data than unknowns.

As an example of this processing, we consider magnetics data from Badlands Bombing Range, South Dakota. This site was used as a bombing and artillery range between 1942 and 1977. The data presented here were acquired with the MTADS system in 1997 over an area contaminated by 250 lb. practice bombs (Barrow and Nelson, 2001). Figure 1 shows a typical fit, and dipole parameters, obtained for a single target using the UXOLab software (Lelièvre et al., 2004).
Figure 1. Observed, predicted and residual data for a single target in the Badlands bombing range magnetics data. The bottom right panel shows the model parameters and estimated uncertainties.

The parameter *Angle* denotes the angle of the dipole moment with the earth’s magnetic field.

Given the model parameters obtained from inversion, we are faced with the problem of deciding whether a target is likely to be a UXO. This problem is commonly approached as a *supervised* learning problem. Supervised classification makes decisions on a *test* data set for which labels, or ground truth, are unknown. The classifier performance is optimized using a *training* data set for which labels are known. In unsupervised classification there is only a test data set: labels are unknown for all observed patterns. Unsupervised classification is often a more difficult problem than supervised classification.

Template matching is a simple solution to supervised learning problems. This approach compares observed data with pre-defined templates and is essentially a pattern matching problem. A particular pattern is classified based upon the template which best matches the observed data. A typical measure of similarity is the correlation between the data and the template. While template matching is an intuitive approach to classification, it is often difficult to generalize templates so that they can match all possible permutations of the data.

A more common approach is to use the model parameters estimated via inversion as basis vectors in an *M*-dimensional *feature space*. The basis vectors in this space are selected so that feature vectors belonging to different classes occupy different regions of the feature space. Figure 2 shows a two-dimensional feature space with training and test data sets for data from Badlands bombing range.

The feature space is spanned by the logarithm of the estimated dipole moment and the logarithm of the angle of the dipole moment with the earth's magnetic field. We choose the strength of dipole moment as a feature because we expect clutter items will tend to have smaller moments than the large UXOs encountered at this site. The motivation for choosing angle as a second feature is that the angle of the induced dipole moment of an ordnance item is constrained to lie within approximately 60 degrees of the earth's magnetic field (Billings, 2004). In contrast, clutter items may have a remanent (permanent) magnetization which can cause the angle of the estimated moment to deviate widely from the earth's field.
direction. For analysis, we use the logarithms of moment and angle, because these quantities are typically log-normally distributed.

The ideal goal of classification is to assign correct labels to the test data, i.e. in the middle panel of Figure 2 we wish to determine whether each item is UXO or clutter. This can be challenging because of potential overlap in feature space that occurs when items from different classes have similar feature values. In the rightmost panel of Figure 2 we have replotted the training data and superposed three test items which need to be classified. Just from visual inspection of the plot, it is likely that items 1 and 2 are respectively clutter and UXO. Classification of item 3 is more difficult since it lies near the boundary of any intuitive region that would separate the two classes. In fact, item 3 is a clutter item. In practical classification, we need to formulate strategies that deal with such overlaps.

Given training and test data sets, the goal of a statistical classifier is to find an optimal partition of the feature space. Within statistical classification there are two approaches to generating a decision rule. A generative algorithm seeks to model the underlying distributions which produced the observed data, often assuming a parametric distribution such as the Gaussian. A discriminative algorithm is not concerned with underlying distributions but rather seeks to identify decision boundaries which provide an optimal separation of classes.

2.1 Generative classifiers

The starting point for any generative classifier is Bayes rule:

$$P(\omega_i|x) \propto P(x|\omega_i)P(\omega_i)$$

The likelihood $P(x|\omega)$ is the probability of observing the feature vector $x$ given the class $\omega$. The prior probability $P(\omega)$ quantifies our expectation of how likely we are to observe class $\omega$ before (i.e. prior to) observing any feature vector data. Bayes rule translates the prior probability into a posterior probability $P(\omega|x)$. The posterior is the probability that we have observed class $\omega$ given the observed feature vector.

Bayes rule provides a mechanism for classifying feature vectors: assign feature vector $x$ to the class with the largest posterior probability. This is optimal in the following sense: if we assign each feature vector to the most probable a posteriori class, then we expect the total number of misclassifications to be minimized.

Application of Bayes rule to classification requires knowledge of the prior probabilities and the form of the likelihood function. In the Bayesian framework, prior distributions play a central role: they quantify our subjective expectations. When Bayes rule is used in the form given above, the prior probabilities weight the relative importance of classes.
The likelihood function can take either a parametric or nonparametric form. The parametric representation assumes a probability distribution for each class and tries to estimate the parameters of these distributions from the training data. The most common parametric classifier is discriminant analysis, which assumes a Gaussian form for the likelihood function. To implement this classifier, we estimate the mean and covariance of each class (UXO or clutter) using the training data. Quadratic discriminant analysis computes a separate covariance for each class. In this case, the decision boundary is a quadratic function in the feature space (see figure 4). Alternatively, if the same class covariance is assumed equal for all classes, then the classifier produces a linear decision boundary in the feature space (linear discriminant analysis).

A parametric classifier which has been used for UXO classification is the Gaussian likelihood ratio. This classifier considers the ratio of posterior probabilities

$$\lambda = \frac{P(x|\omega_1)P(\omega_1)}{P(x|\omega_2)P(\omega_2)}$$

so that $\lambda = 1$ corresponds to a feature vector $x$ on the decision boundary between classes $\omega_1$ and $\omega_2$. This classifier is a reformulation of discriminant analysis (either linear or quadratic) and produces the same decision list.

Assuming a parametric form for the likelihood function greatly simplifies the problem of estimating class distributions. However, this assumption may be difficult to justify if limited training data are available. In this situation, we may turn to nonparametric methods, which define a likelihood function directly from the training data. A representative nonparametric classifier which has been applied to UXO discrimination is the probabilistic neural network (PNN). This classifier represents the class distributions as superpositions of Gaussian “kernels”, with each kernel centered on a feature vector in the training data.

The class distributions estimated with a PNN can produce a highly nonlinear decision boundary in the feature space. The complexity of the decision boundary is governed by the width of the kernel functions: broad kernel functions produce smooth distributions and approximately linear decision boundaries. Narrow kernel functions, on the other hand, can produce complex decision boundaries. Estimating the required classifier complexity is a regularization problem: we must balance our desire to “fit” the training data with the need to correctly classify an unseen test data set. Optimal regularization parameters (e.g. kernel smoothings) are typically estimated using the available information in the training data.

2.2 Discriminative classifiers

Instead of estimating posterior probability distributions, discriminative classifiers directly define a decision boundary to classify test data. Finding a decision boundary which separates the training data and generalizes well to the test data can be approached as a constrained optimization problem. A commonly used classifier of this form is the support vector machine (SVM). The basic idea of this classifier is illustrated in figure 3. We maximize the margin between classes, subject to the constraint that the training data are classified correctly. The margin is defined as the perpendicular distance between support planes. As shown in figure 3, a support plane is a line (or plane in higher dimensions) such that all feature vectors in a class fall to one side of that line.
Figure 3. Support vector machine formulation for linearly separable feature data belonging to two classes (squares and circles). The classifier tries to maximize the margin between support planes, subject to the constraint that the feature data are correctly classified.

A more general formulation of the support vector machine allows for nonlinear decision boundaries with overlapping classes. The idea is to map the feature data to a higher-dimensional space where the training data become separable. We then construct the decision boundary in this space (Hastie et al., 2001). As with the PNN, support vector machines have adjustable parameters which control the complexity of the resulting decision boundary.

3. Application of statistical classifiers to data from Badlands bombing range

Figure 4 shows the decision boundaries for the several classifiers applied to feature data from Badlands bombing range. The logarithms of the dipole moment and the angle of the moment with the earth's magnetic field are used as features. The training data in this example are a random sample of 10% of all available data.

To compute the decision boundary for a generative classifier, we estimate the class posterior probabilities for all classes at a grid of points in the feature space. By retaining the maximum posterior probability (evaluated over all classes) at each point, we generate an image of the classifier output. Minima of this image then correspond to decision boundaries in the feature space.

To validate the performance of our statistical classifiers on the test data, we use the receiver operating characteristic (ROC) curve. This curve shows the proportion of UXO items which have been found as a function of the proportion of clutter items which have been found. To generate this curve, we threshold on the output of each classifier. This thresholding sweeps a decision surface through the feature space until all UXO items in the test data have been identified. Animations of the thresholding process in this feature space can be viewed at http://www.eos.ubc.ca/~lberan/research.html.
Figure 4. Classifier solutions for the Badlands bombing range data. The feature space is spanned by strength of the estimated dipole moment and the angle of the moment with the earth’s magnetic field. Squares are UXO items and triangles are clutter items. Grayscale images for the first three classifiers show the maximum posterior probability evaluated over all possible classes. Decision boundaries correspond to minima of this function. Squares show the maximum posterior probability evaluated over all possible classes. Decision boundaries correspond to minima of this function. The rightmost plots show the receiver operating characteristic (ROC) curve generated by thresholding on the output of each classifier.

a. Linear Discriminant Analysis
b. Quadratic Discriminant Analysis
c. Probabilistic Neural Network
d. Support Vector Machine
4. Discussion

This article is not an exhaustive survey of the pattern recognition algorithms which can be used for UXO discrimination. However, our hope is that the preceding discussion can provide the reader a basic familiarity with the common approaches to statistical classification.

The statistical classifiers described here are not the only available approach for discriminating between UXO and clutter items. A commonly-used discrimination method is to dig every anomaly with an amplitude above a specified threshold. While simple thresholding circumvents the sometimes laborious process of inversion, anomaly amplitude is not a particularly robust feature for discriminating between UXO and clutter. For example, UXO items at depth may produce a comparable anomaly amplitude to shallower clutter items. Feature extraction via inversion can provide more useful diagnostics for discrimination. In cases where multiple data types are available, we can threshold on some linear combination of the available features. At a higher level of sophistication, we may take nonlinear combinations of features.

It should be emphasized that increasing classifier sophistication does not necessarily imply better performance. An advantage of simple thresholding or “rule-based” classifiers is that they do not require any information regarding the distributions of UXOs and clutter at a field site. In contrast, statistical classifiers model the decision boundary between UXO and clutter and so we need information about both classes before we can make predictions for the test data. Also, statistical classifiers can perform poorly if the initial training data set is not representative of the test set. On the other hand, statistical classifiers can learn from the training data as digging proceeds and thereby improve their performance. It may be necessary to use a combination of approaches (statistical and rule-based) to obtain the desired performance at a field site.

Initial studies carried out by Beran (2005) suggest that this may be a reasonable strategy.

A weakness of statistical classifiers is their dependence upon a representative training data set. The rarity of UXOs at some field sites can make random sampling an inefficient way to generate training data: we are unlikely to obtain enough UXO ground truth to model the UXO class with confidence. Recent work by Zhang et al. (2004) has focused upon generating the initial training data set from an unlabelled feature space. This is an unsupervised learning problem: we must identify the classes without any information about UXO and clutter in the feature space. Initial results with this approach have significantly improved performance compared to random sampling.

Current discrimination practices at actual field sites lag far behind the latest research. While acquisition of digital geophysical data is now the norm, most stakeholders/regulators require excavation of all targets with anomaly amplitudes up to a predefined threshold. Future efforts in UXO discrimination research should therefore focus on practical application of more sophisticated rule-based and statistical classifiers. Currently, there are several efforts in this direction that we, and others in the field, are undertaking. These include (i) large scale test and evaluation of discrimination performance at live sites; (ii) development of user-friendly software, such as the recent implementation of the statistical classification toolbox within the UXOLab software package; and (iii) engagement with stakeholders and regulators at live-sites so that discrimination methodologies can be demystified and ultimately accepted as part of the clearance process.

References


