IDENTIFICATION OF UNEXPLODED ORDNANCE FROM CLUTTER USING NEURAL NEWORKS

Anna Szidarovszky, Zonge Engineering and Research Organization, Tucson, AZ
Mary Poulton, University of Arizona, Department of Mining Engineering, Tucson, AZ
Scott C. MacInnes, Zonge Engineering and Research Organization, Soldotna, AK

Abstract

The largest costs associated with subsurface Unexploded Ordnance (UXO) remediation are associated with removing non-UXO. Discrimination between UXO and non-UXO is important for both cost and safety reasons. A neural network was developed to distinguish between UXO and non-UXO clutter using TEM data. There are two stages for the learning process of neural network, training and validation. A synthetic dataset was created using actual acquisition configurations, with varying amounts of random noise. This dataset included 934 UXO targets representing 7 different UXO types, and 789 clutter objects based on four templates with varying size and random asymmetry. The results show 97% accuracy for correctly classifying clutter, and 97% accuracy for correctly classifying UXO.
The level of success for classification is based on the classification Receiver Operating Characteristic (ROC) curves. The ROC curve represents the relationship between UXO classified correctly (Hit rate) versus clutter miss classified (False alarm).

Introduction

In many countries around the world, UXO remain from warfare or from military practice. UXO pose a problem to society in two main ways: cost of detection and removal, and safety of civilians and removal crews. UXO are differentiated from landmines and this paper focuses solely on UXO.

The United States Department of Defense estimates that around 15 million acres at approximately 1500 sites are contaminated with unexploded ordnance (UXO) in the United States (Bowers, Bidwell, 1999). UXO are explosive ordnance which have been primed, fused, armed or otherwise prepared for action, and which have been fired, dropped, launched, projected or placed in such a manner as to constitute a hazard to operations, installations, personnel or material and remain unexploded either by malfunction or design or for any other cause (Department of Defense 1997). UXO must be located, and fully cleared from the area, before land use is possible. The process of detecting and removing UXO is not yet cost and time efficient. It is estimated by the United States Army Corps of Engineers that 75% of the money spent on UXO removal projects is spent on removing non-UXO or "false positives" (Bowers, Bidwell, 1999). Newly developed technology that would minimize the excavation of false positives would help reduce these costs.

One of the methods that is widely used is the "mag and flag" technique; a slow, labor intensive method where operators use hand-held metal detectors to locate potential UXO, then place a flag on the spot. With this method one can only detect a buried ferrous object, and differentiation between clutter and UXO is seldom possible. A second widely used method is the Time Domain Electromagnetic Method (TEM), where a transmitter produces a square wave current, which transmitted into a transmitter coil. TEM has less depth of investigation than the magnetic method, since the signal falls of more rapidly. UXO usually stay at surface, or get buried no more than 8 meters. (Leonard R. Pasion and Douglas W. Oldenburg August 2001). As a trade off, the TEM method is less sensitive to geologic noise and it can detect non-ferrous objects. A synthetic dataset was used for this project since the field equipment was under development. Seven different UXO types and four fragment pieces were modeled using the program DNTDipole developed by Zonge Engineering over a wide size range and scaled with a random symmetry.

Introduction to Neural Networks

Engineering decision making and design are usually based on mathematical models that represent or compute quantitative relations between the model parameters. These parameters are physical values, decision variables, and certain quantities showing the consequences of our decisions. Most relations representing physical interrelations or consequences are either based on experimental observations or on theoretical facts. In both cases these relations are uncertain or inaccurate, since measured values have errors and theoretical considerations are always based on simplifying assumptions. The complexity and diversity of mathematical model types make it impossible to develop a general model, which can be applied in most cases. Nature however has provided such a tool, the human nervous system in the brain, which can transfer complicated signals and information. The brain has evolved to process complex signals, summarize them and make intelligent decisions. Biological neural networks are trained during a lifetime based on repeated experiences, and they are not controlled by complicated mathematical models, only by the data resulting from repeated experiences. This system has been very successful in biological systems and the idea of using the same approach for computer-based pattern recognition has been developed over the past 60 years. The fundamental idea of biological neural networks as pattern recognition devices can be traced back to the late 19^{th} century to the noted psychologist William James (1890). He discussed memory functions including understandability, predictability and the simple structure of human brain and neural system. He also pointed out that the brain does not necessarily think abstractly, it is constructed to survive. Therefore it has ability that a good engineer should possess: solve the problem as well as possible and as simply as possible (Anderson and Rosenfeld, 1988).

The architecture of an ANN has three components: input neurons, output neurons and hidden neurons (Poulton, 2001). The structure of a multi-layer perceptron (MLP) is shown in figure 1, where *m*

is the number of inputs, n is the number of outputs and p is the number of nodes in the hidden layer.

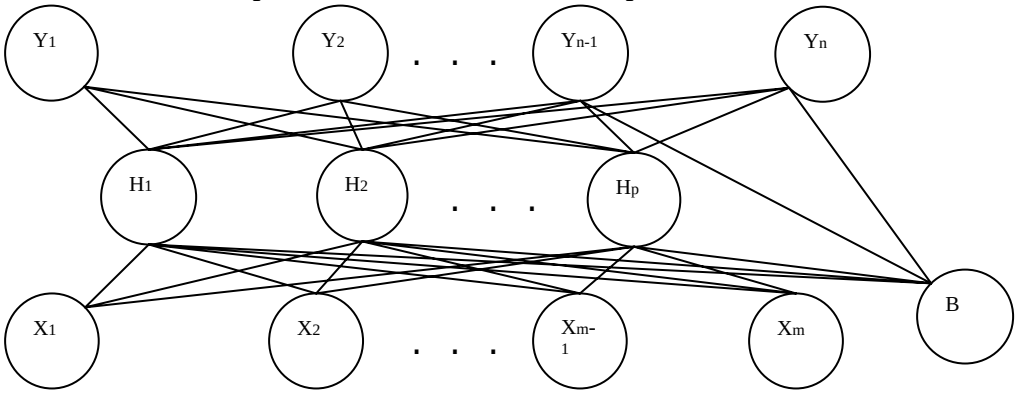

Figure 1: Multi-layer Perceptron structure

We note, that the structure shows that every node in the hidden layer is connected to all nodes in the input and output layers. The bias unit (B) assumes unit input values and its role is to speed up the learning process. The bias unit is connected to all hidden nodes as well as to all outputs. Each node receives several inputs, depending on the location of this node in the hierarchy its inputs can be original inputs or hidden variables. It also generates an output which is directed toward to the nodes of the next hidden layer or to the output variables of the system. For notational convenience let $u_1,...,u_k$ denote the inputs and v denote the output of a MLP. The value received by the MLP is a linear combination of its inputs:

$$Sum = \sum_i W_i U_i + W \,, \tag{1}$$

where the coefficients W_i are the multiples of the inputs and the W is the coefficient of the bias unit (with unit input value). The output is then obtained from sum via an Activation function:

$$V = f(Sum). \tag{2}$$

In practical applications f is selected either the sigmoid , $f(Sum) = \frac{1}{1+e^{-Sum}}$ (3)

or the Hyperbolic tangent function

$$\bar{f}(Sum) = \frac{e^{Sum} - e^{-Sum}}{e^{Sum} + e^{-Sum}} \tag{4}$$

.

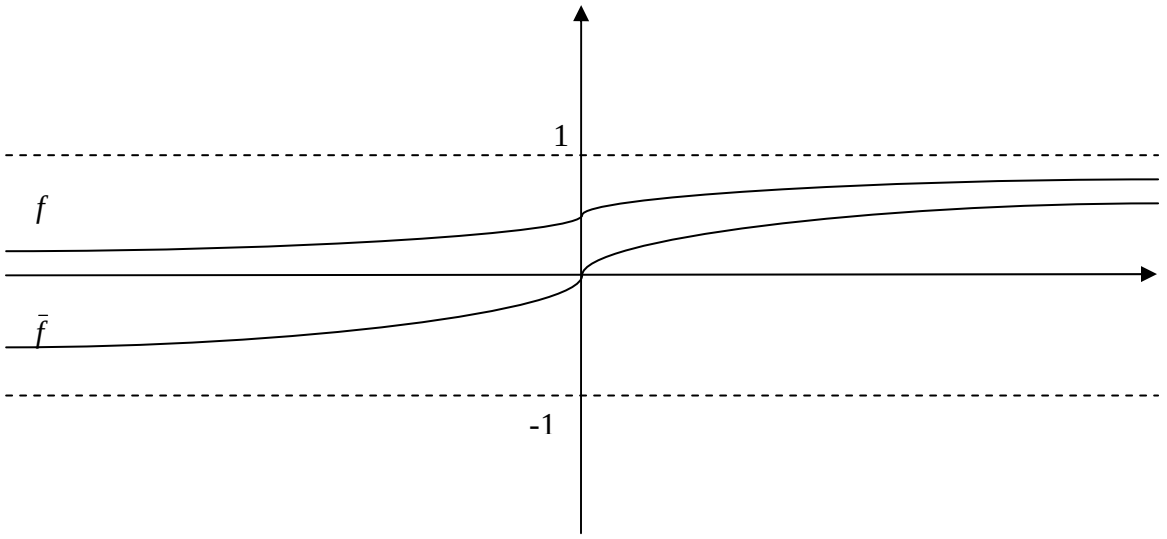

Figure 2: Graph of the f Sigmoid and $\bar{f}$ Hyperbolic tangent activation functions

Observe that f transforms Sum into the unit interval (0,1), and $\bar{f}$ transforms it into (-1,1). Simple algebra shows that $\frac{1}{2}\left(1+\bar{f}(Sum)\right)=\frac{1}{2}\left(1+\frac{e^{Sum}-e^{-Sum}}{e^{Sum}+e^{-Sum}}\right)=\frac{e^{Sum}}{e^{Sum}+e^{-Sum}}=\frac{1}{1+e^{-2Sum}}=f(2Sum)$ (5)

implying, that the two most popular activation functions are basically the same, they are linear transformations of each other. Methods for determining the connection weights between the nodes will be discussed in the next section. With initial random weights the first hidden variables can be computed from the inputs by using equations (1) and (2), and then the second hidden variables can be obtained, and so on in the case of multiple hidden layers, and finally all outputs can be obtained. That is, with given set of weight values W_{ij}, the output Y_k can be directly and very easily obtained for any input selection $x_1,...,x_m$.

This recursive procedure is a function $\underline{y}=\underline{F}(\underline{x},\underline{w})$, where $\underline{x}$ is an input vector, $\underline{w}$ is the vector having the connection weights in its components and $\underline{y}$ is the output vector.

Algorithms for computing the error and updating connections weights are called learning algorithms. Training and validation are the two stages of any learning process. First about half of the data are used to determine the connection weights and then the second half of the data are used to check the accuracy of the obtained input-output relation $\underline{F}$.

Let T denote the number of input-output data being used in training, let $\underline{x}^{(t)},\underline{y}^{(t)}$ denote the $t^{\underline{th}}$ $(1\le t\le T)$ data pair. The fit of the input-output relation $\underline{F}$ can be characterized by the least-squares type measure

$$Q=\sum_{t=1}^{T}\sum_{k=1}^{n}\left[F_k\left(\underline{x}^{(t)},\underline{w}\right)-y_k{}^{(t)}\right]^2, \tag{6}$$

where F_k and y_k denote the $k^{\underline{th}}$ component of $\underline{F}$ and $\underline{y}$.

Notice, that this measure depends on only the weight selection $\underline{w}$, therefore the best fitting neural network is obtained by minimizing function Q with respect to the unknowns $\underline{w}$. In the mathematical optimization literature there are many different methods to be used here, therefore there are several training methods being used in practical applications.

The actual threshold for deleting unimportant inputs, and the final neural network structure are based on the subjective judgment of the user.

For this project a synthetic dataset was generated. The original model parameters for the synthetic dataset for the seven different UXO types are based on free-air measurements made by Jonathan Miller of AETC for SERDP project UX-1313 (Miller, J.T., 2002). Where it was necessary, the size of the training sample was increased by duplicating the properties of the measured objects. For modeling "irregular", non-UXO items, the properties measured on four fragment pieces over a wide size range was scaled and random asymmetry was added.

Each training object instance was given a random depth and orientation and then synthetic TEM data were calculated. The TEM data represented results from multi-loop array configurations used during static follow-up surveys at potential UXO locations selected during a dynamic UXO detection survey. Each transient has 31 time windows spaced logarithmically between 0.04 and 24 msec. The synthetic data included additive Gaussian noise with moderate amplitude. Field data quality may well be of lower quality.

Introduction to modeling program, DNT-Dipole

Target characterization is generated by inversion of multi-component TEM transient data to three polarizability transients, one for each axis of an anisotropic dipolar target model. The shape of each polarizability transient is parameterized in several ways. Integration of polarizability transient dp(t)/dt with respect to time generates a polarizability amplitude, p0, which measures the overall polarizability magnitude. The shape of the transient is described with the Pasion-Oldenburg [1] model $dp(t)/dt = k \cdot \exp(-t/\tau) \cdot (a+t)^{-b}$. Ratios of polarizability magnitude and Pasion-Oldenburg model parameters provide measures of target symmetry.

Target model parameter estimates from the TEM inversion program are used as a basis for target classification by the ANN.

One of the ANN training challenges was reducing the large number of possible input parameters to the most diagnostic set. In our case we had 133 different parameters to choose from. In the application of neural nets too many parameters slow the computation and create erroneous patterns; too few will not give the best result. Histogram analysis on 30 parameters from the given data allowed us to reduce them to 21 (most correlated with the input and least correlated with each other).We also scaled the parameter values by the equation $P_{scaled} = \log(P + const)$. Then we created training and testing sets with 21 inputs, and one output. The 21 input parameters are the following:

P0_R longitudinal/transverse polarizability P0_1/P0_T

P0_E polarizability eccentricity = (P0_2 + P0_3)/P0_1

P1_R longitudinal/transverse polarizability P1_1/P1_T

P1_E polarizability eccentricity = (P1_2 + P1_3)/P1_1

P1 weights late-time more heavily than does P0, so it is a little noisier but if the data quality is good may express object symmetry better than P0. The effect of permeable object elongation becomes more apparent in mid to late-time data.

AVG(PB) average Pasion-model b (unitless)

PB_T transverse Pasion-model b (unitless)

PB_R longitudinal/transverse Pasion-model b

PB_E Pasion-model b eccentricity

PB_1 target u-axis Pasion-model power-law decay, b (unitless exponent)

PB_2 target v-axis Pasion-model power-law decay, b (unitless exponent)

PB_3 target w-axis Pasion-model power-law decay, b (unitless exponent) Pasion-Oldenburg model parameter PB indicates mid-time dP/dt slope and should be a good indicator of an object with the expected conductivity/relative permeability properties. (see Pasion and Oldenburg, 2001 for more detail)

RMS(P0) root mean square polarizability amplitude (cm^3)

P0_T transverse polarizability = (P0_2 + P0_3)/2 (cm^3)

P0_1 target u-axis polarizability (cm^3)

P0_2 target v-axis polarizability (cm^3)

P0_3 target w-axis polarizability (cm^3)

Inversion of TEM dB/dt transients to a triaxial dipole model generates a dP/dt polarizability transient for each target-model axis. In addition to parameterizing polarizability transient shape with the Passion-Oldenburg model (Pasion,L.R., and Oldenburg, D.W., 2001), DNTDipole calculates two polarizability magnitude parameters by integrating dP/dt:

$$P0 = \int_0^\infty \frac{dP(t)}{dt} \cdot dt \quad \text{and} \quad P1 = 1000 \cdot \int_0^\infty \frac{dP(t)}{dt} \cdot t \cdot dt \tag{7}$$

where $\frac{dP(t)}{dt}$ is in $cm^3/usec$, P0 is in cm^3 and P1 is in $msec \cdot cm^3$

P0 characterizes overall polarizability magnitude, while the first moment P1 emphasizes late-time behavior where the character of target shape is more well developed. P0 and P1 values for each target axis are used in derived parameters like “polarizability eccentricity” = (P0_2 + P0_3)/P0_1 that measure the target’s symmetry (Bell, T.H., Barrow, J., Miller, J.T., 2001). P0 is a good measure of object electrical size, but UXO can vary over such a large size range that measures of size won't be of much help unless a site happens to have a very limited selection of UXO types.

RMS(P1) polarizability first-moment RMS amplitude (msec*cm^3)

P1_T transverse polarizability first-moment = (P1_2 + P1_3)/2 (msec*cm^3)

P1_1 target u-axis polarizability (msec*cm^3)

P1_2 target v-axis polarizability (msec*cm^3)

P1_3 target w-axis polarizability (msec*cm^3) (Bell, Barrow, Miller, 2001

The output for the ANN is one processing element that calculates a value of 0 for clutter and 1 for UXO.

The 7 different UXO types used are:

60 mm mortars (676 examples)

81mm mortars (137)

20mm (23)

30mm (25)

37mm (25)

105mm (24)

155mm (24)

A total of 934 UXO targets were modeled.

Scaled properties of four clutter items over wide range of size and added random asymmetry were also generated for a total of 789 clutter objects.

Results and discussions

The MLP architecture consisted of 21 inputs, a single hidden layer with 9 nodes, and a single output that computed a value of 1 for UXO and 0 for clutter.

25 UXO were misclassified as clutter (output<0.5) out of 934. (5 - 20mm; 3 - 30 mm; 15 - 60 mm; and 2 - 81mm were missed). The accuracy of the classification for UXO was 97.3%. For the clutter, 19 objects were misclassified as UXO (output >0.5) out of 789 yielding a classification accuracy of 97.6% for clutter.

Only 5 UXO and 3 clutter items were missed classified if the threshold was set to be 0.3 instead of 0.5 The Receiver Operating Characteristic (ROC) curve approach is used to assess the quality of the classifications. The ROC curve represents the relationship between UXO classified correctly (Hit rate) versus clutter misclassified (False alarm) (see Figure 3).

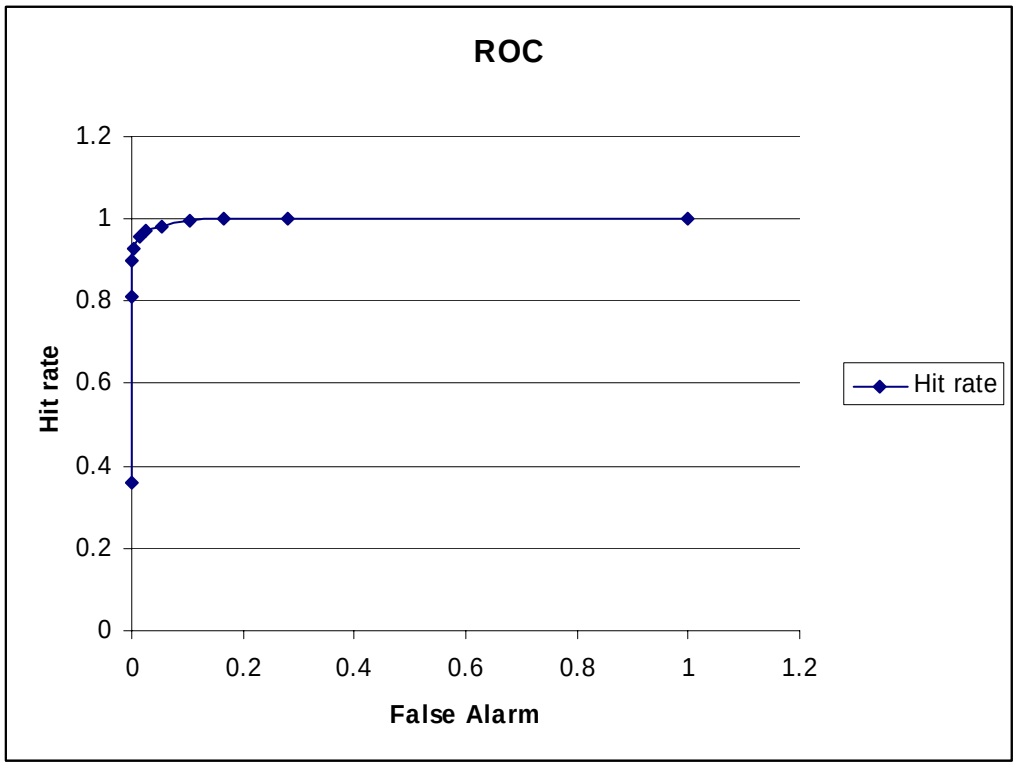

Figure 3: The Receiver Operator Characteristic (ROC) Curve examines the trade off in false alarms and hit rates as the threshold for correct classification is changed. Curves that rise sharply along the y-axis and then become abruptly horizontal near a y-value of 1 indicate a robust classification.

We examined the data parameters to try to determine the reasons for the misclassifications. One of the input parameters which seems to play an important role is the parameter P0(E), eccentricity of the polarizability.

The following 2 figures (Figure 4a and 4b) compare the parameter values of the missed UXO (error for the ANN output has to be greater than 0.5 to be classified as missed) versus the seen UXO and clutter (error for the ANN output is less than 0.1 for these cases).

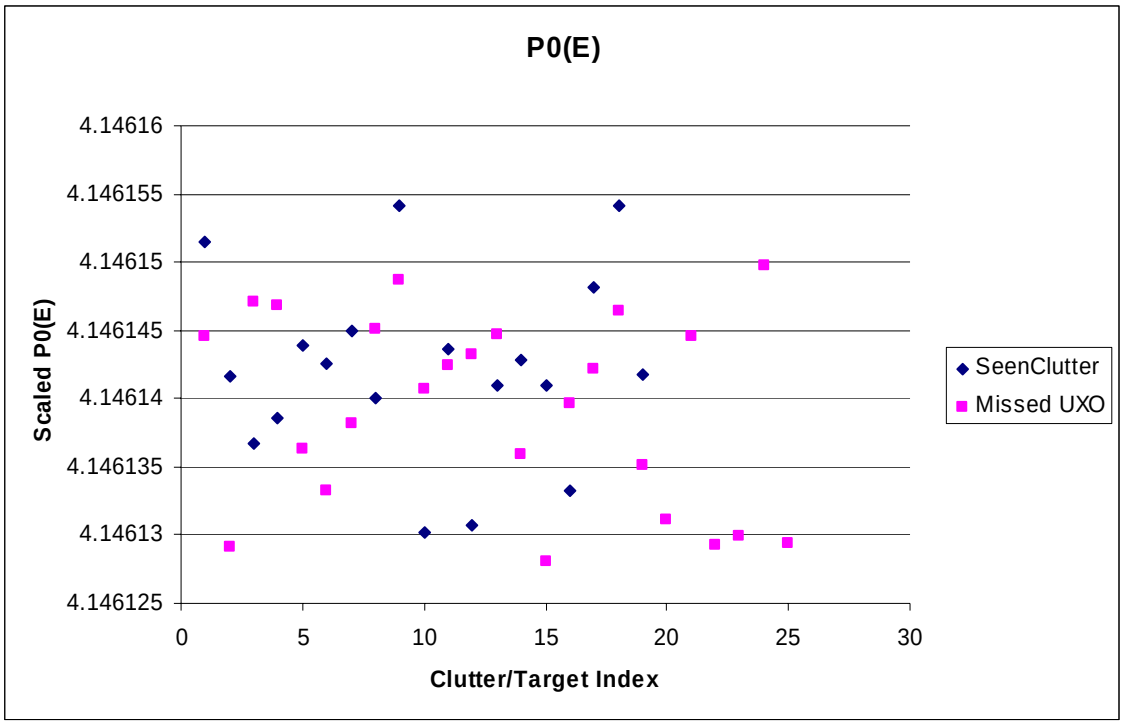

Figure 4a: The values for parameter P0(E) for the missed UXO are similar to the clutter which indicates for these particular UXO and clutter, the parameter does not have any discriminating capability.

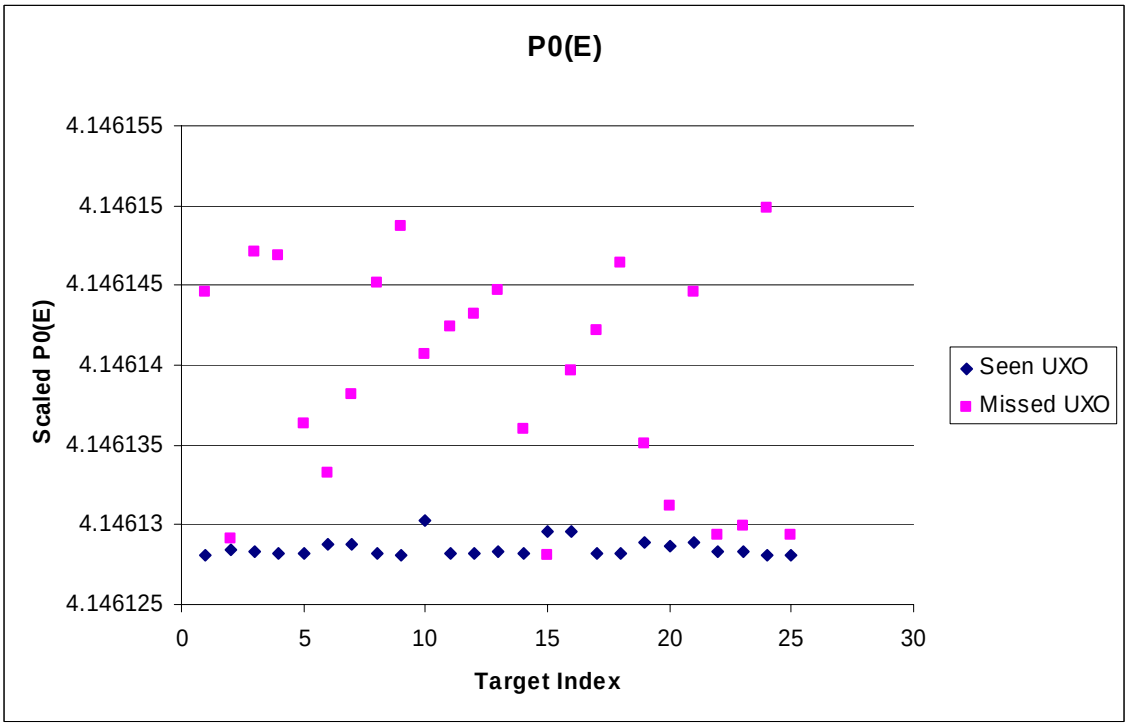

Figure 4b: The values for the parameter P0(E) for correctly classified UXO show a consistent range of values while those for the missed UXO show a wide range of scatter indicating these particular UXO have a pattern that is inconsistent with most of the other UXO.

Misclassification by the ANN is due to the similarity of some of these parameters for the UXO and clutter and we must error on the side of classifying clutter as UXO rather than allowing UXO to go undetected.

There are other parameters as well that are responsible for good classifications. Figure 5 shows how shallower objects are more difficult to classify correctly. This plot shows the relationship between the depth of burial and the error between the actual output and the desired output.

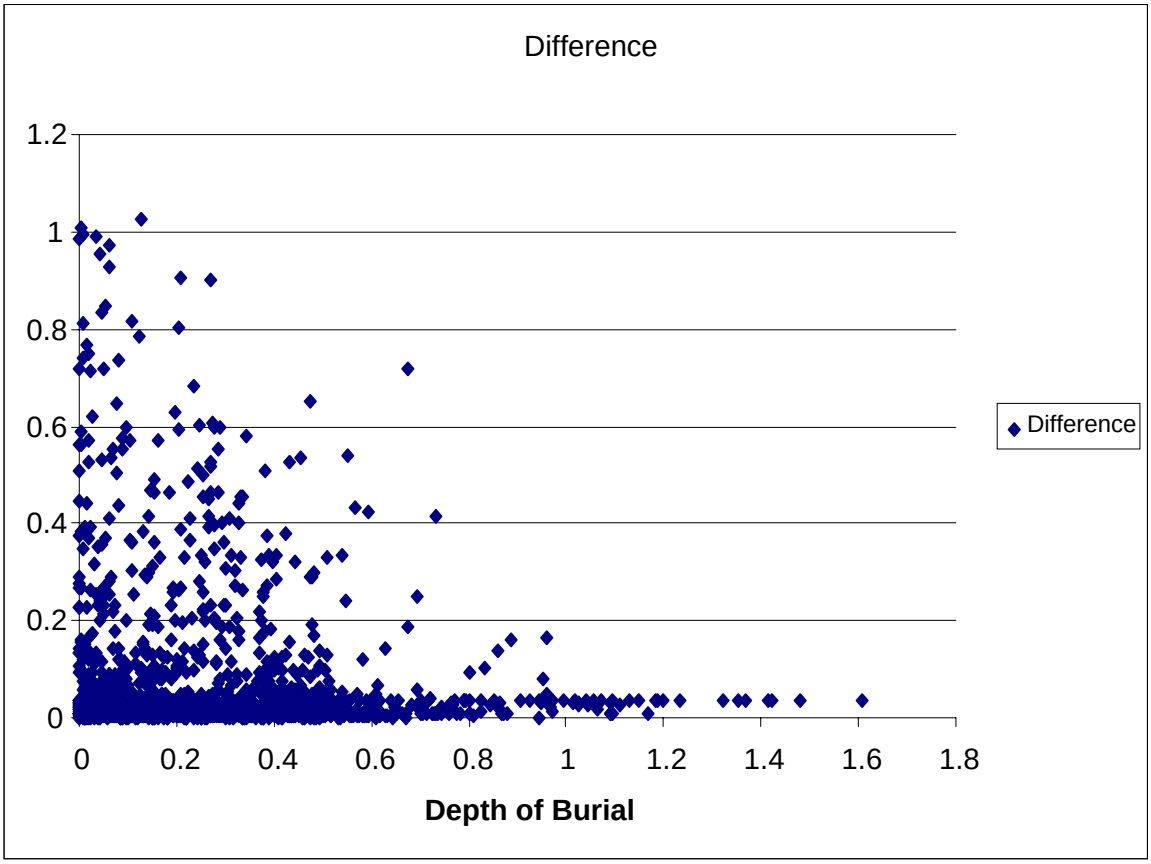

Figure 5: The depth of burial of UXO and clutter plotted against the amount of error in the classification (the difference between the desired value of 0 for clutter or 1 for UXO and the value calculated by the ANN). A value of 1 on the y-axis would indicate a complete misclassification; a value of 0.5 on the y-axis indicates an ambivalent classification. Lower difference values indicate a successful classification.

Conclusion

In many countries around the world, UXO pose a problem to society in the form of cost of detection and removal, and safety of civilians and removal crews. Finding a good method for discriminating between clutter and UXO is very important. Using ANN, this discrimination is possible for synthetic data with 97 % accuracy. In order to increase the accuracy of correctly classifying all UXO, we need to have a good understanding of the important parameters, and to understand how they are responsible for the outcome. These parameters include the ones DNTDipole provides, along with depth of burial, mass, length, width, orientation. Once we have that information, the next step is to create a dataset, which will be optimized for classifying all UXO and minimizing the number of the

misclassified clutter. When it is impossible to distinguish between UXO and clutter in terms of the physical properties we have to call the object UXO for safety reasons. When the field system is fully functional, synthetic training data can be supplemented or replaced with field data to allow the system to learn from field experience.

References

Anderson, J., and E. Rosenfeld, 1988, Neurocomputing: Foundations of Research: MIT Press, Boston, MA.

Bell, T., B. Barrow, and J. Miller, 2001, Subsurface discrimination using EM induction sensors: IEEE Tran. on Geoscience and Remote Sensing, v39, p1286-1293.

Bowers, J., and B. Bidwell, 1999, Geophysics and UXO detection: The Leading Edge (December 1999),1389-1391.

James, William, 1890, Psychology (Briefer Course): Holt, New York.

Miller, J.T., 2002, Quantitification of UXO variability for target discrimination, SERDP Project UX-1313

Pasion, L., and D. Oldenburg, 2001, A Discrimination Algorithm for UXO Using Time Domain Electromagnetics. Jour. Eng. & Envir. Geophys., vol. 28, p. 91-102.

Pasion, L., Billings, S., and Oldenburg, D., 2001, UXO Discrimination Using Time Domain Electromagnetic Induction: UXO Counter-Mine Forum, 2-9.

Pasion, L., and D. Oldenburg, 1999, Locating and Determining Dimensionality of UXOs Using Time Domain Electromagnetic Induction. in *SAGEEP 1999*. Oakland, CA.

Poulton, M., (Ed.), 2001, Computational Neural Networks for Geophysical Data Processing: Pergamon, Amsterdam.

United States Department of Defense. “DOD Ammunition and Explosives Safety Standards.” DOD 6055.9-STD August 1997. 2004.<https://www.denix.osd.mil/denix/Public/ES-Programs/Explosives/Safety/cover.html