Using Machine Learning to Complement and Extend the Accuracy of UXO Discrimination Beyond the Best Reported Results of the Jefferson Proving Ground Technology Demonstration

Larry M. Deschaine (SAIC & Chalmers Univ. of Tech); Richard A. Hoover, Joseph N. Skibinski, and Janardan J. Patel (SAIC); Frank D. Francone (Chalmers Univ. of Tech. & Register Machine Learning Technologies); and M. J. Ades

Keywords: Unexploded ordnance, anomaly detection, geophysics, genetic programming

Abstract

The accurate discrimination of unexploded ordnance from geophysical signals is very difficult. Research has demonstrated that using a machine learning technique known as linear genetic programming in concert with human expertise can extend the accuracy of unexploded ordnance discrimination past currently published results. This paper describes how linear genetic programming offers the promise of creating real-time unexploded ordnance discrimination.

THE UXO CHALLENGE

When military munitions do not function as intended or fully detonate, they become unexploded ordnance (UXO). Many challenges arise if and when the UXO is found on land destined for use for something other than military testing or training. [Van Antwerp 2001]. It has been estimated that more than 16 million acres of land on closed, transferred, and transferring ranges are potentially contaminated with UXO [GAO 2001]. The estimated cost for remediating the U.S. training ranges alone is at least $14 billion, and this number is likely understated [GAO 2001]. In aggregate, UXO poses a potentially significant problem in the United States as former military lands are used for other purposes. The scale grows rapidly if battlefields and minefields located worldwide are included in this estimate.

Fortunately, it is relatively easy to identify UXO using a variety of geophysical sensors. Available sensors include magnetometers; radar; and active electromagnetic, chemical, and electro-optical sensors. Unfortunately, separating a UXO anomaly in geophysical data from noise anomalies or scrap metal originating from successfully detonated munitions is more difficult. This difficulty results in a large number of “false alarms” that require dangerous (and expensive) exposure of the feature, and physical examination to determine if the source of the feature is scrap or UXO. As many as 75% of all anomalies investigated are false alarms such as scrap metal, ordnance debris, cans, and wire [Butler et al., 2001]. Active research continues on a number of fronts to enhance the quality of data collected and the methods of data analysis.

Among the contributing difficulties are the size, shape, and composition of the UXO. The size can vary from small arms munitions (such as a 37-millimeter-diameter projectile) to large-target, airborne-delivered munitions (such as a 2,000-pound bomb that is 477 millimeters in diameter and 1.74 meters long). Shape can vary from spherical (a projected grenade is 40 millimeters in diameter, and a hand grenade is 64 millimeters in diameter) to linear (a 4.5-inch rocket is 114 millimeters in diameter and 1.9 meters long). As for composition, beyond a variety of explosives that carry differing chemical signatures, composition can range from ferrous iron (which is magnetic) in the 2,000-pound bomb to generally aluminum (which is nonferrous) in rocket bodies.

CURRENT EVALUATION TRENDS AND LIMITATIONS

It has been determined that geophysical surveys (a) can detect UXO within definable limits, (b) cannot effectively discriminate UXO anomalies from “false alarm” anomalies, and (c) cannot always identify UXO [Butler et al., 2001]. The challenge exists, therefore, to more effectively use the geophysical data collected. The means to evaluate the geophysical data traditionally fall into two general areas: signature matching and analytical modeling.

Signature matching is a technique that has had fair success. In theory, a particular bomb with a specific weight and shape should have a unique signature. In practice, however, a number of things will alter the geophysical signature that is measured. If carefully performed geophysical measurements are compared to a database of historical measurements that have been exposed and identified, a match of “signatures” and ordnance may be developed [Damarla and Ressler 2000]. This approach requires that a significant body of information be gathered in a consistent fashion for comparison and evaluation.

Beyond these requirements, some specific physical limitations also constrain this approach. The physical characteristics of the soil at a particular site may mask the parameter being measured, resulting in an overall reduced
signature; therefore, the approach is site- and even soil-specific. In addition, for a particular ferrous metal target, the orientation of the UXO in the earth (north to south versus east to west) will present significantly different measurements. Add the third dimension, and any change in the attitude of the feature from flat lying to vertical can significantly alter the measured response. If we consider that the remnant magnetization of the iron body is dependent upon the metallurgical properties and the thermal, mechanical, and magnetic history of the specimen [Breiner 1973] and also consider the thermal effects due to frictional heating that occurs when a bomb passes through soil, we can see that a broad variety of signatures may result from a single type of target, with a single orientation, within a single soil type.

The modeling approach commonly focuses on the response of munitions to particular geophysical sensors. The model-based approach may be either exact or use an approximate forward-modeling algorithm to determine the set of model parameters needed to replicate the measured responses and relating the model parameters to physical parameters or targets [Khadr et al., 1998]. The variables associated with remnant magnetization described in the preceding paragraph are effectively focused to identify a limited class of ordnance. This kind of approach can be general as applied to a particular sensor [Norton and Witten 1998]. Recent advances related to time-domain electromagnetic methods are summarized and advanced in Pasion and Oldenburg 2001. There is little information in the published literature describing integrated sensor analysis for a more comprehensive interpretation of the geophysical data.

It is the integration of various sensors as well as expert-derived and machine-learning-derived approaches that we are exploring as part of this research and development (R&D) work. The integration algorithm we chose to integrate this information is linear genetic programming (LGP). Field deployment optimization is accomplished using evolutionary strategies.

GENETIC PROGRAMMING

Good, detailed treatments of GP may be found in Banzhaf et al., 1998 and Koza et al., 1999. In brief, the particular type of GP used in this work is LGP. The LGP algorithm is conceptually surprisingly simple. It starts with a population of randomly generated computer programs. These programs are the “primordial soup” on which computerized evolution operates. Then, LGP conducts a “tournament” by selecting four programs from the population—also at random—and measures how well each of the four programs performs the task designated by the LGP developer. The two programs that perform the task best “win” the tournament.

The LGP algorithm then copies the two winner programs and transforms these copies into two new programs through crossover and mutation transformation operators. In short, the winners have “children.” These two new child programs are then inserted into the population of programs, replacing the two “loser” programs from the tournament. LGP repeats these simple steps over and over until it has written a program that performs the selected task.

LGP creates its “child” programs by transforming the tournament-winning programs. The transformations used are inspired by biology. For example, the LGP mutation operator transforms a tournament winner by changing it randomly. For instance, the mutation operator might change an addition instruction in a tournament winner to a multiplication instruction. Likewise, the LGP crossover operator causes instructions from the two tournament-winning programs to be swapped—in essence, an exchange of genetic material between the winners. LGP crossover is inspired by the exchange of genetic material that occurs in sexual reproduction in biology.

LINEAR GENETIC PROGRAMMING USING DIRECT MANIPULATION OF BINARY MACHINE CODE

Machine-code-based LGP is the direct evolution of binary machine code through GP techniques [Nordin 1994; Nordin and Banzhaf 1995a; Nordin and Banzhaf 1995b; Nordin et al., 1998; Nordin 1999]. An evolved LGP program, therefore, is a sequence of binary machine instructions. For example, an evolved LGP program might be comprised of a sequence of four, 32-bit machine instructions. When executed, those four instructions would cause the central processing unit (CPU) to perform operations on the CPU’s hardware registers. Below is an example of a
simple, four-instruction LGP program that uses three hardware registers.

\[
\begin{align*}
\text{register 2} &= \text{register 1} + \text{register 2} \\
\text{register 3} &= \text{register 1} - 64 \\
\text{register 3} &= \text{register 2} \times \text{register 3} \\
\text{register 3} &= \text{register 2} / \text{register 3}
\end{align*}
\]

While LGP programs are apparently very simple, it is actually possible to evolve functions of great complexity using only simple arithmetic functions on a register machine [Nordin and Banzhaf 1995b; Nordin et al., 1998].

After completing a machine-code LGP project, the LGP software decompiles the best evolved models from machine code into Java, ANSI C, or Intel Assembler programs [Register Machine Learning Technologies 2002]. The resulting decompiled code may be linked to the optimizer and compiled, or it may be compiled into a DLL or COM object and called from the optimization routines.

The linear machine code approach to GP has been documented to be between 60 and 200 times faster than comparable interpreting systems [Fukunaga et al., 1998; Nordin 1994; Nordin et al., 1998]. This speed allows for more extensive search of the solution space, resulting in models of higher accuracy [Deschaine and Francone 2002].

WHY MACHINE-CODE-BASED LINEAR GENETIC PROGRAMMING?

At first glance it is not at all obvious that machine-code-based LGP is a strong candidate for the modeling algorithm of choice for the UXO discrimination challenge. The problem is complex and multidimensional, but over the past 3 years, a series of tests has been performed on both synthetic and industrial data sets. LGP has consistently produced very good results on a variety of data sets when compared to other machine-learning techniques [Deschaine and Francone 2002].

In brief, the machine-code-based LGP software has become our modeling tool of choice for addressing the UXO challenge for several reasons.

- Its speed permits us to conduct many runs in realistic timeframes on a desktop or multi-CPU computer, resulting in consistent, high-precision models.
- The LGP algorithm is well designed to prevent overfitting and to produce robust solutions.
- The models produced by the LGP software execute very quickly when called by an optimizer.
- Tests show that solutions exceeding the best published UXO discrimination results have been achieved.
- The solutions are directly interpretable by an expert geophysicist to ensure they make sense. For example, LGP directly derived Darcy’s Law from sparse noisy data, and the output was clearly \(Q=KIA\).

UXO DISCRIMINATION AND LINEAR GENETIC PROGRAMMING

The U.S. Department of Defense (DOD) has been responsible for conducting UXO investigations at many locations around the world. These investigations have resulted in the collection of extraordinary amounts of geophysical data with the goal of identifying buried UXO.

Evaluation of UXO/non-UXO data is time-consuming and costly. The standard outcome of these types of evaluations is maps showing the location of geophysical anomalies. In general, what these anomalies might be (e.g., UXO, non-UXO, boulders) cannot be determined without excavation at the location of the anomaly.

To test whether this LGP approach was viable, we analyzed publicly available data from the Jefferson Proving Ground (JPG) Technology Demonstration Program, which are posted on the Joint UXO Coordination Office Web page. DOD made this information publicly available so that anyone could use the data to develop UXO discrimination algorithms.

The JPG test site consisted of 160 buried anomalies. Some were UXO, while others were non-UXO. The experiment was designed to test various geophysical instruments and analysis algorithms to see if any one method was significantly better than the others.

To conduct a fair comparison test of our analysis versus those that took part in the actual technology demonstration project, we neither contacted the people in charge of the test site nor the vendors who collected the information. The analyst had no insight into the data or guidance about the project other than the posted raw data sets and the public report. Furthermore, the public report, which contained information on how the other algorithms performed on the test plot, was not reviewed by the analyst until the initial solution had been produced.

The geophysical data sets available to the public for independent UXO-discrimination-algorithm development fully disclose which targets are UXO and which are not. To assess the validity of the algorithm we had developed, we randomly segregated the data into three sets to create a blind data set.
The test protocol was as follows: to determine the viability of using LGP, we first placed information about the 160 anomalies in the LGP algorithm and ran various tests. The LGP was able to perfectly discriminate between UXO and non-UXO. Because we left no data out to test the algorithm, the results indicated only that the LGP technique might work. That is, we now had information it was possible that a signal was present in the data that was useful for discrimination between UXO and non-UXO. This type of test is not valid for the final solution, as the algorithm may simply have been memorizing, instead of generalizing, the data set.

To develop an algorithm that will repeatedly predict UXO versus non-UXO with known accuracy, the algorithm must perform well on multiple data sets, including a blind data set. The JPG test site data sets contained a small number of examples, only 160 data points. To create our data sets for this analysis, we randomly divided the data set into three sets: 50 points for training, 50 for validation, and 60 for an applied data blind test. Of these, there were 110 non-UXO anomalies and 50 UXO targets in the data set. The applied data set had 15 UXO targets and 45 non-UXO anomalies.

Figure 1 shows the performance of the published results from ten analyses conducted by vendors who provide UXO services as part of the JPG Phase IV project [Jefferson Proving Ground 1999]. The horizontal axis shows the performance of each algorithm in correctly identifying anomalies that did not contain buried UXO, whereas the vertical axis shows the performance of each algorithm in correctly identifying anomalies that did contain buried UXO. The angled line in Figure 1 represents what would be expected from random guessing.

![Figure 1. LGP UXO discrimination solution compared to results from the JPG Phase IV UXO Discrimination Project](image)

Figure 1 points out the difficulty of modeling these data. Most algorithms did little better than random guessing; however, the LGP algorithm derived a best-know model for correctly identifying UXO and for correctly rejecting non-UXO using various data set configurations.

The gray dot in the upper right-hand corner of the figure shows the LGP solution on the unseen data. Because the number of data points was small, we used resampling techniques to estimate the 95% confidence interval on this solution. The black rectangle in Figure 1 shows that interval.

The JPG Phase IV data set is considered an easier data set to succeed on than some because the type of scrap metal used to seed the non-UXO targets was not derived from UXO. It is easier to discriminate this metal from UXO-derived metals. The JPG Phase V test plot uses UXO-derived scrap metal. The authors will analyze these data when they become available.

These results or conclusions are not meant to discredit the other analyses or vendors in any way. Working with these data over the past 8 months, testing various combinations of inputs, data preconditioning, and the like, we have concluded that the UXO problem is indeed an extremely difficult one to solve. What we have demonstrated is that the information content from the geophysical sensors combined with advanced machine-learning techniques is sufficient to develop high-accuracy UXO discrimination algorithms. We have accomplished what we have set out to do. An advantage of this approach is that it provides a confidence estimate of whether a target is UXO or non-UXO, so the site-specific remedial action plan, known as the dig sheet, can be prioritized. It also identifies which sensor, or combination of sensors, provides input valuable to UXO discrimination such that data collection tasks can also be optimized.

The learning occurred on the training data set. The best-evolved programs were selected using the training and the validation data set. Whether the “true” structure of the UXO discrimination solution was captured was again measured by using the applied data set. In other words, the applied data (also referred to as testing data) played no part in training or in best-program selection. Accordingly, the results on the applied data measured how well the evolved solution generalized to unseen data. These are the results, the unseen data results, shown in Figure 1.

The accuracy of the field-deployed tool is expected to be within the area of this rectangle. This approach allows for the dynamic incorporation of results from field investigation. It is expected to become more accurate as more data are used, and the spread of the confidence intervals will narrow as more data and field-result (ground truth information) become available for analysis.

Hence, the approach is consistent on projects in which a test plot is constructed with about a dozen buried ordinance items about which everything is known and for which the results are expected to be translatable to masses of data about which nothing is known. At first the confidence
interval will be wide, but the tool will become increasingly more accurate as deployment occurs.

EVOLUTION STRATEGIES OPTIMIZATION

Evolution Strategies (ES) was first developed in Germany in the 1960s. It is a very powerful, general-purpose parameter-optimization technique [Rechenberg 1994; Schwefel 1995; Schwefel and Rudolph 1995]. Although we refer in this work to ES, it is closely related to Fogel’s Evolutionary Programming (EP) [Bäck and Schwefel 1993; Fogel 1992]. Our discussion here applies equally to ES and EP. For ease of reference, we will use the term ES to refer to both approaches.

ES uses a population-based learning algorithm. Each generation of possible solutions is formed by mutating and recombining the best members of the previous generation. ES pioneered the use of evolvable “strategy parameters.” Strategy parameters control the learning process; therefore, ES evolves both the parameters to be optimized and the parameters that control the optimization [Banzhaf et al., 1998].

ES has the following desirable characteristics for use in our methodology:

- ES can optimize the parameters of arbitrary functions. It does not need to be able to calculate derivatives of the function to be optimized, nor does the researcher need to assume differentiability and numerical accuracy. Instead, ES gathers gradient information about the function by sampling [Hansen and Ostermeier 2001].
- A substantial amount of literature over many years has demonstrated that ES can solve a very wide range of optimization problems with minimal customization [Rechenberg 1994; Schwefel 1995; Schwefel and Rudolph 1995; Hansen and Ostermeier 2001].

Although very powerful and not prone to getting stuck in local optima, typical ES systems can be very time-consuming for significant optimization problems. Canonical ES, therefore, often fails the requirement of efficient optimization.

But in the past 5 years, ES has been extended using the ES-CDSA technique [Hansen and Ostermeier 2001]. ES-CDSA allows a much more efficient evolution of the strategy parameters and cumulates gradient information over many generations rather than a single generation as in traditional ES.

As a rule of thumb, where \( n \) is the number of parameters to be optimized, users should allow between 100 and 200\((n+3)^2\) function evaluations to get optimal use from this algorithm [Hansen and Ostermeier 2001]. While this rate represents a large improvement over previous ES approaches, it can still require many calls by the optimizer to the model to be optimized to produce results. It is still very important, therefore, to couple ES-CDSA with fast-executing models, which is where the LGP solution becomes important because it executes in a fraction of a second on a standard PC. This combination opens up the possibility of real-time, fast, optimal UXO discrimination.

OPTIMIZING THE LGP-DERIVED UXO MODELS

As discussed above, the problem of UXO affects millions of acres worldwide and includes both training areas and former battlefields. The estimated cost for remediating the U.S. training ranges alone is at least $14 billion, and this number is probably understated, particularly if worldwide estimates are included. A very real cost of cleanup (or non-cleanup) could be the injury or death of people.

Currently, too large a portion of the resources available for responding to UXO challenges is expended on digging up sites where UXOs are expected, but which turn out to be false alarms—that is, false positives. Expenditures on this activity result in limited funding being available for remediation of genuine UXOs. Machine-code-based LGP has derived the most accurate UXO discriminator among published results to date [Jefferson Proving Ground 1999] by a wide margin. This LGP UXO/non-UXO identification success opens up the assessment and optimization of response to the UXO issue on both the program and the project level. Some examples of the possibilities include the following:

- The presence or absence of UXO can be assessed using remote, nondestructive technology such as land- or air-based sensors, including geophysics and various wavelength sensors. When the sensor technology and data collection have been developed to a high degree of efficiency and accuracy, wide areas can be screened and analyzed to reduce the footprint of regions needing further investigation. This feature will help manage the sheer size of the challenge.
- Areas of interest identified as requiring further investigation can be prioritized and ranked using good information on the probability or absence of UXO. This ranking will integrate the LGP UXO solution with multicriteria/multiobjective decision support models.
- Site-specific remedial action plans (or dig sheets) can be optimally designed to focus efforts on high-probability UXO-containing areas. When a decreased predicted likelihood of UXO presence and a field-verified
absence of UXO are demonstrated, a stopping point for remedial activities, based on scientific principals and field validation, will be provided.

SUMMARY AND CONCLUSIONS

The work was conducted as an R&D effort by the authors to establish whether sufficient information is contained in geophysical sensor signals to develop high-accuracy UXO discrimination algorithms. We have concluded that high-accuracy UXO/ non-UXO discrimination is achievable.

We have completed the R&D phase of this work and are in the early stages of building a comprehensive, integrated modeling and optimization system to handle complex UXO discrimination challenges. We believe a combination of machine-code-based LGP (for modeling) and ES-CDSA (for optimization) provides the best blend of available tools and algorithms for this task.

The results described in this paper have been independently peer reviewed and approved. To gain industry acceptance, the results of these tools must be demonstrated independently and in more realistic (i.e., impact area as opposed to test plot) settings before LGP and ES-CDSA can be used on a widespread scale in the UXO remediation industry. The Environmental Security Technology Certification Program (ESTCP) was established by DOD to demonstrate and validate technologies that target urgent environmental needs and apply DOD-wide. For these reasons, the authors of this paper will seek independent scientific endorsement by conducting one or several demonstrations through ESTCP or similarly functioning organizations. Future versions of this paper will include the results of those demonstrations.

AUTHOR CONTACT INFORMATION

Larry M. Deschaine (SAIC & Chalmers University of Technology)
Science Applications International Corporation 360 Bay Street, Suite 200
Augusta, GA 30901
(706) 724-5589, ext. 227
Larry.M.Deschaine@alum.mit.edu

Richard A. Hoover
Science Applications International Corporation 6310 Allentown Boulevard
Harrisburg, PA 17112
(717) 901-8835
Richard.A.Hoover@saic.com

Joseph N. Skibinski
Science Applications International Corporation 11251 Roger Bacon Drive, Mail Stop R-3-3
Reston, VA 20190
(703) 810-8994
Joseph.N.Shibinski@saic.com

Janardan J. Patel
Science Applications International Corporation 11251 Roger Bacon Drive, Mail Stop LOC 641-R3
Reston, VA 20190
(703) 318-4267
Janardan.J.Patel@saic.com

Frank D. Francone [RMLT & Chalmers University of Technology]
Register Machine Learning Technologies, Inc.
11757 Ken Caryl Avenue, #F512
Littleton, CO 80127
(720) 981-8710
ffrancone@aimlearning.com

Maurice J. Ades, Ph.D.
Consultant
mjades@msn.com

REFERENCES


