Discrimination of Unexploded Ordnance from Clutter Using Linear Genetic Programming

Frank D. Francone[1], Larry M. Deschaine[2], Tom Battenhouse[3], Jeffrey J. Warren[4],

[1] RML Technologies, Inc and Chalmers University of Technology, 7606 S. Newland St., Littleton CO USA 80128
ffrancone@aimlearning.com
[2] Science Applications International Corp. and Chalmers University of Technology, 360 Bay St., Augusta GA USA 30901
Larry.M.Decshaine@SAIC.com
[3] Science Applications International Corp, 360 Bay St., Augusta GA USA 30901
Thomas.R.Battenhouse.Jr@SAIC.com
[4] Science Applications International Corp., 6310 Allentown Boulevard., Harrisburg, PA USA 17112
Jeffrey.J.Warren@SAIC.com

Abstract. We used Linear Genetic Programming (LGP) to study the extent to which automated learning techniques may be used to improve Unexploded Ordinance (UXO) discrimination from Protem-47 and Geonics EM61 non-invasive electromagnetic sensors. We conclude that: (1) Even after geophysicists have analyzed the EM61 signals and ranked anomalies in order of the likelihood that each comprises UXO, our LGP tool was able to substantially improve the discrimination of UXO from scrap—preexisting techniques require digging 62% more holes to locate all UXO on a range than do LGP derived models; (2) LGP can improve discrimination even though trained on a very small number of examples of UXO; and (3) LGP can improve UXO discrimination on data sets that contain a high-level of noise and little preprocessing.

1 Introduction

The Department of Defense ('DoD') recently stated: "The UXO cleanup problem is a very large-scale undertaking involving 10 million acres of land at some 1400 sites." [1] One of the key problems is, according to DoD, ". . . instruments that can detect the buried UXO's also detect numerous scrap metal objects and other artifacts, which leads to an enormous amount of expensive digging. Typically 100 holes may be dug before a real UXO is unearthed!"[1] Buried UXO poses a hazard to life-and-limb and further prevents huge tracts of land—frequently urban—from being returned to civilian use.

Geophysicists have recently begun gathering magnetic and electro-magnetic data about potential UXO sites using non-invasive, above-ground sensors. They gather UXO data by pulling various active and passive sensors across a UXO site and record the sensor readings. This process is called Digital Geophysical Mapping ('DGM'). Unfortunately, the digital signal for UXO frequently resembles the signal from clutter (scrap metal that poses no danger to the public) and OE fragments (pieces of UXO that have sheared-off during impact). Figure 1 illustrates the difficulty of distinguishing UXO from clutter. Currently, most UXO discrimination from DGM is made by human experts analyzing the DGM signal.

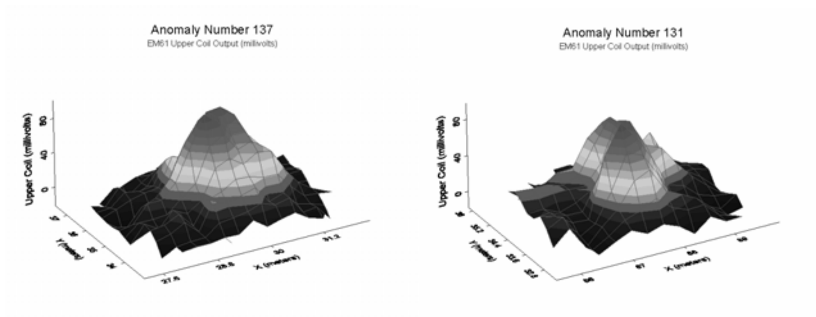

Fig. 1. Signature of buried UXO (example, left) versus clutter (example, right)

This paper reports the successful application of a process we refer to as UXO/MineFinder™ service to the problem of UXO discrimination on two data-sets acquired from DoD UXO test-beds. This is a multi-step process that includes five high-level tasks:

1. **Acquisition** of DGM data by geophysicists;[1]
2. **Anomaly Identification** by geophysicists of physical locations where the DGM indicates may be potential UXO;
3. **Extraction** of relevant features pertaining to each anomaly by geophysicists;
4. **Ranking** of anomalies by the likelihood that the anomalies are UXO using the Linear Genetic Programming [2,3] software, Discipulus™ [4], and;

[1] We studied DGM data from the Jefferson Proving Grounds IV [5] and V [6] test-plots (JPG-IV and JGP-V, respectively) for the two different phases of this study. For this study, DGM data acquisition (Step 1) was performed by third-party contractors engaged by the DoD. In particular, we used data acquired by NAVEA on a Protem-47 from JPG-IV [7] and by the National Research Laboratory ('NRL') on an EM61 for JPG-V [8].

5. **Characterization** of UXO (such as ordnance type, depth, and orientation).

This paper focuses on Step 4, the discrimination process, and is organized as follows.

- **First**, Linear Genetic Programming is at the heart of our process. We will briefly describe the LGP algorithm and software used in this study in Section 3 below.
- **Second**, Phase I of this study was a prove-out of the discrimination portion of our process on the Jefferson Proving Grounds IV data from NAVEA. Section 4 will discuss the methodology we used for this Phase I, the results obtained, and compare those results with the results obtained by other contractors.
- **Third**, Phase II of this study was completed in February of 2004. Phase II tested Steps 2-4 above—anomaly identification, feature extraction and LGP ranking of anomalies on the Jefferson Proving Grounds V data from the NRL. Section 5, below, discusses the methodology we used for Phase II, our results, and compares them with the best-known results from other contractors.

3 Linear Genetic Programming

Linear Genetic Programming ('LGP') is at the core of our process. We used Discipulus™, which is a Machine-Code-Based, Multi-Run, Linear GP system. This automated learning software distinguishes our process from other UXO discrimination techniques, which are based mostly on human engineering expertise.

3.1 Genetic Programming

Genetic Programming (GP) is the automatic, computerized creation of computer programs to perform a selected task using Darwinian natural selection. GP developers give their computers examples of how they want the computer to perform a task. Here, the 'examples' would be paired inputs and outputs—the inputs being features of the DGM and the output representing ground-truth, that is, is the anomaly a UXO? From these examples, GP software then writes a computer program that performs the task described by the examples. Good overall treatments of Genetic Programming may be found in [2,9].

LGP represents the evolving population of programs as linear genomes—that is, a linear strings of executable instructions to the computer [10]. The LGP algorithm is surprisingly simple. A detailed description of it is available in [4, 3].

Machine-code-based, LGP is the direct evolution of binary machine code through GP techniques [10,11]. Thus, an evolved LGP program is a sequence of binary machine instructions. 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 [10,12]. The machine-code approach to GP has been documented to be between 60 and 200 times faster than comparable interpreting systems [10,11,13].

Multi-Run LGP is based on our observation that, if one performs many runs with the same parameters, varying only the random seed, a histogram of the results from the different runs will usually describe a normal distribution with a tail of good solutions extending to the right [3,14]. To know that the full extent of the distribution of runs has been discovered, it is necessary to perform multiple LGP runs until a stable distribution is achieved. The LGP software we used performs this process automatically [4].

After completing a multi-run LGP project, the LGP software decompiles the best evolved models from machine code into Java, ANSI C, or Intel Assembler programs [4]. The resulting decompiled code may be linked to other code and compiled or it may be compiled into a DLL or COM object.

Having now described the LGP software used, we will now turn to describing, in order, the two phases of this applied LGP project.

4 Phase-One: Proof-of-Concept Study of JPG-IV, Protem-47 UXO DGM Signatures

Phase I of this investigation was a proof-of-concept phase that applied LGP to the JPG-IV test-bed data. JPG-IV is a research quality test-bed. UXO and clutter were buried at known locations and depths. Contractors with sensors were invited to measure the geophysical signatures at these known locations [5].[2] Altogether, sensor readings for 50 UXO and 110 clutter items were available from the JPG-IV site.

From the DGM, contractors attempted to discriminate between UXO and clutter [5]. The DGM acquired by the various sensors at the JPG-IV locations were then made available to other contractors to test their ability to discriminate between UXO and clutter and it is these data that were used in Phase 1 of this study. We used the data collected by NAEVA on the JPG-IV site using a Protem-47 transmitter and receiver, configured with twenty time-gates [7]. The data from all twenty time-gates were made available as inputs to the LGP algorithm.

The data were randomly split into training and validation sets, which were used, respectively to train the LGP algorithm and to select the best programs for testing on unseen data. A portion of the data was held-back from the training and validation sets. LGP was run until a stable distribution of results was produced. At that point, the best program

[2] This technique of gathering data is significantly different than is typical on an actual UXO site. On an actual UXO site, there is no preexisting knowledge of where to look for UXO. Accordingly, DGM must often be conducted for the entire site. Thus, the JPG-IV data is very high-quality data gathered from known anomalies and using sensors in a stationary mode, rather than being pulled across the site.

produced by LGP on the training and validation data sets was selected as the best-program from the project.

Once a best-program was produced by LGP, it was tested on the held-out data. All results reported here are on the unseen, held-out data.

The LGP software produced excellent results on the NAEVA data [15]. As noted above, out of ten contractors, only one produced results that were better than random guessing [5]. Their results are shown as small black points on Figure 2. Our results are shown as a large black point in the upper right-hand-corner of Figure 2. The arrow represents the amount by which our approach improved the discrimination results obtained by NAEVA using the same data we used. The difference between our results and those of the next best contractor, Geophex, Ltd., were statistically significant at the 95% level.

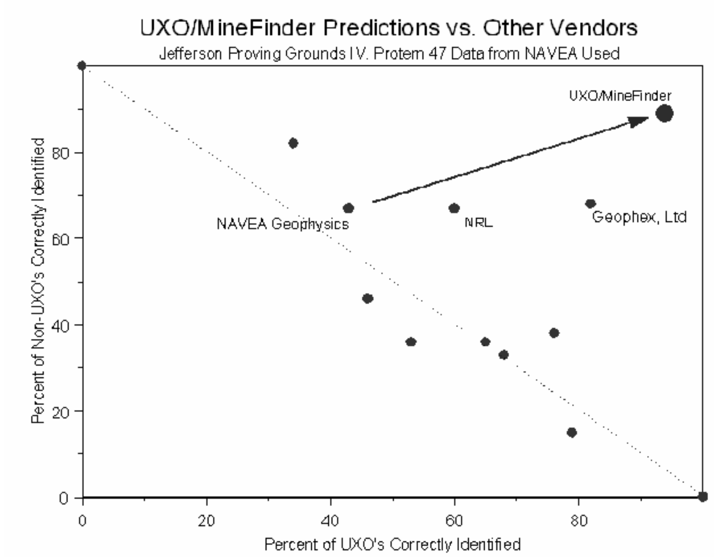

Fig. 2. Comparison of UXO/MineFinder UXO Discrimination Results and Other Vendor's Discrimination Results on JPG-IV Test-Bed Data

4.4 Conclusion

This test established that using LGP as a classifier tool for UXO discrimination was very promising. Accordingly, further testing was required to prove-out our process as an integrated production service. The next section details our findings in that regard.

5 Phase II: Production Prove-out on the JPG-V, EM61 UXO DGM Signatures

Our Phase II prove-out was performed to test our process on production grade data where it was necessary to integrate data-cleansing, anomaly identification, feature extraction and selection and UXO discrimination into a single package. This section reports our methodology and results for that prove-out.

5.1 Data Used in Production Prove-Out

We selected the NRL data from Jefferson Proving Grounds V, Area 3 [6] as being most suitable to the goals of this project because:

- The JPG-V project was designed to mimic an actual impact area. The DoD's JPG IV project, which failed to do so in several regards [5];
- The JPG-V data was from production-quality instruments and collection techniques, rather than research-quality;
- The JPG-V data was gathered by contractors in a manner consistent with data acquisition in the field—trailers bearing sensors were pulled across the JPG-V site.
- The NRL data appeared to be the cleanest data available.

From the various data feeds collected by the NRL, we chose the NRL's single time-channel time-domain electromagnetic induction sensor data (MTADS), collected in Area 3 of the JPG V demonstration survey. The instrument used to collect the data was an EM61 [8].

5.2 Preprocessing Applied to NRL Data

While the NRL data appeared to be the highest-quality data amongst the three contractors, no calibration data was available from the NRL to iron out inconsistencies from track-to-track. On examination of the NRL data, there appeared to be substantial calibration problems as among tracks. In addition, the background level of geomagnetic noise varies substantially within single tracks of data. We elected not to try to correct the calibration problems and background noise level problems; rather, we decided to allow the LGP classifier to model the calibration and background noise along with the target signals.

Our preprocessing was, therefore, limited to gridding the data using standard procedures recommended by the Geosoft Oasis-Montaj geophysical software (an industry-standard in geophysical surveying) for target identification using the default parameters.

5.3 Anomaly Identification

Anomaly selection represents the first critical UXO screening step. Advanced geophysical data processing attempts to balance target area selection of UXO with weak observed signals (because background clutter or nearby UXO create a complex signal) with the selection of a disproportionate number of target areas containing no UXO.

We used Geosoft Oasis-Montaj to select potential targets in the JPG-V, Area 3 field. The procedure was straightforward. We set a threshold of six millivolts as the smallest anomaly that should be identified as a target. Given that threshold, Geosoft located three-hundred forty-two anomalies that we thereafter treated as our targets for classification.

5.4 Feature Extraction for the Identified Targets

The JPG-V Area 3 data from NRL was transformed into a set of 1D (point statistics) and 2D (spatial statistics) features. Only physically meaningful features were generated so that the physical interpretation of evolved prediction algorithms was not prohibitively difficult.

The 1D features used were the Geosoft created values for Upper and Lower Coil readings for each identified target.

Generation of 2D features included analysis of both the gridded data and the raw data. 2D analyses of gridded data utilized standard image processing algorithms. Techniques, such as subsampling, morphological processing, and 2D filtering, were used to preprocess the gridded data. An example of extracted 2D features are the major and minor axes of an anomaly at a point 50% of the way up the anomaly and at a point located 95% of the way toward the bottom of the anomaly from the top.

5.5 Methodology for Creating LGP Target Rankings

In UXO cleanup, the primary tool used to guide engineers is called a 'dig-list.' It identifies each anomaly and its coordinates. A dig-list is often prioritized. That is, it includes instructions where to dig first, where to dig next and so forth.

This project was posed to create an efficient prioritization for the JPG-V site dig list. Efficiency is tested by how many holes must be dug (starting with the highest ranked hole and proceeding down the list) until all UXO have been located. The fewer holes dug before all UXO are located, the lower the cost of the project [17]. This measure of performance is used in preference to a classic machine learning classification confusion matrix approach because this methodology was used by the DoD in assessing contractor's performance on the JGP-V test bed [6].

Our principal concern about the JPG-V, Area 3 data we used was that Geosoft located only nineteen UXO and thirty-three OE fragments.[3] This is a very small number of positive examples of UXO. Many of our decisions in configuring LGP for this project were directed at minimizing overfitting arising from such a small data set.

There were several sub-tasks performed in deriving anomaly rankings using LGP. They were: (1) Feature selection; (2) LGP Configuration; (3) Creating multiple data sets; (4) Setting LGP parameters; and (6) Converting LGP outputs into Rankings. Each of these steps is discussed below.

Feature Selection. We started with thirty-six features for each anomaly. Given the small number of UXO and fragment signatures, we were confident that we would not be successful with LGP using all of these features as inputs because of overfitting problems. Thus, we went thru a three-step winnowing process to select the most promising features.

The first step of the winnowing process involved statistical analysis of the various features to select those features with the most significant relationship with the classification task and with the lowest cross-correlation amongst the inputs themselves [16]. We used primarily correlation analysis and ANOVA for this step.

The second step involved using the feature set in traditional modeling tools such as logistic regression and classification trees for two purposes: (1) To determine which features provided the most UXO discrimination ability; and (2) To determine whether either of these traditional tools produced satisfactory discrimination results. There were no surprises from this process in terms of feature selection—it merely confirmed our earlier statistical analysis. This step also made clear that these traditional modeling tools did not perform particularly well in discriminating UXO from clutter. Accordingly, we determined that a more powerful modeling tool, such as LGP, was required.

The third step involved further narrowing the number of features used by conducting multiple LGP runs and examining the "Input Impacts" report generated by the LGP software. That report tells which inputs to LGP were actually used by LGP in a significant way to solve the problem [4].

When these three winnowing steps were concluded, we selected eight inputs to LGP for the remainder of our runs.

[3] Altogether, there were twenty UXO's on site. But Geosoft failed to identify one of them as a target. So information about that UXO was never presented to the LGP algorithm.

LGP Configuration. Based on an input-by-input statistical analysis, we determined that it might be possible to use the OE Fragment data points as 'quasi-positive' examples of UXO. ANOVA for many of the extracted features revealed that the mean of their values for OE Fragments was between the mean values for UXO and Clutter. Furthermore, the mean value of those features for fragments was considerably closer to the mean UXO value than the mean clutter value. This raised the possibility that the OE Fragment anomalies contained useful information about what UXO looked like. Because of the small data set size, this possibility was very attractive because it increased the amount of information available to the LGP algorithm about the characteristics of a UXO as opposed to clutter.

Of course, to use OE Fragments in this manner required that we configure LGP for regression and assign different, but sensible, target values for UXO, OE Fragments and clutter. Based on these observations, we configured LGP for regression and we assigned the following values to as the target output to be approximated: For clutter, we assigned a regression target output value of 0. For OE Fragments, we assigned a regression target output of 0.75. Finally for UXO, we assigned a regression target output of 1.0. These values reflected the reality that OE Fragment feature values tended to fall between UXO and clutter feature values but be closer to the UXO feature value than to the clutter value.

Multiple Data Sets. Because there were a total of fifty-two UXO and OE Fragment items, we created fifty-two separate data-sets. Each of those data-sets held-out as unseen data only one of the UXO or Fragment items together with one-hundred-forty-five clutter points for model validation. The clutter points were chosen randomly for each of the fifty-two data sets. After creating the held-out data set, the remaining data points were used for model creation.

LGP Parameter Settings. Several runs were performed on several of the data sets to come up with a parameterization of LGP that provided enough robustness and generalization to solve the problem but not so much as to overfit the data. Satisfactory base settings were derived. LGP was then run separately on all fifty-two data sets using the base parameter settings derived above. Each run was observed while in progress for overfitting—sampling noise makes it unlikely that the same parameters will be optimal for reducing overfitting for all data sets. We took as evidence of overfitting, a clear observed pattern where the fitness on the targets used for training LGP was negatively correlated with the fitness on the held-out targets.

Fewer than half of the runs showed signs of overfitting. For those runs, we progressively changed the LGP parameters so as to reduce the computational power available to the LGP algorithm until observed overfitting was minimized. At that point, we inserted a new random seed into the LGP algorithm and ran it at those parameters. The resulting run was then accepted as the production run.

Converting LGP Outputs into Anomaly Rankings. We converted LGP outputs on unseen data points into anomaly rankings as follows: for each of the fifty-two data sets, the anomalies held out as unseen data were ranked so that the anomaly with the highest LGP output was ranked number 1, the next highest ranked as number 2 and so forth. Then those rankings were averaged for each anomaly over each of the data sets in which the anomaly appeared as an unseen data point. That average ranking was the ranking assigned to a particular anomaly for our simulated prioritized dig list.

5.6 Evaluation of LGP Prioritized Dig-List

The LGP produced rankings of the 342 anomalies in JPG-V, Area 3 were evaluated against UXO predictions on these same data derived from best-known conventional methods. Those best-known results are reported in the DoD's JPG-V final report for Area 3 [6]. The results of the comparison may be stated simply: the previous best UXO discrimination results on these data were reported by the NRL. NRL's rankings of anomalies required that ninety-six holes be dug before the last UXO was located. The LGP prioritized dig list required that only sixty-four holes be dug before the last UXO was located. Thus, the NRL ranking required digging 62% more holes than did the LGP based ranking. Figure 3 shows the results of our rankings in a pseudo-ROC format.

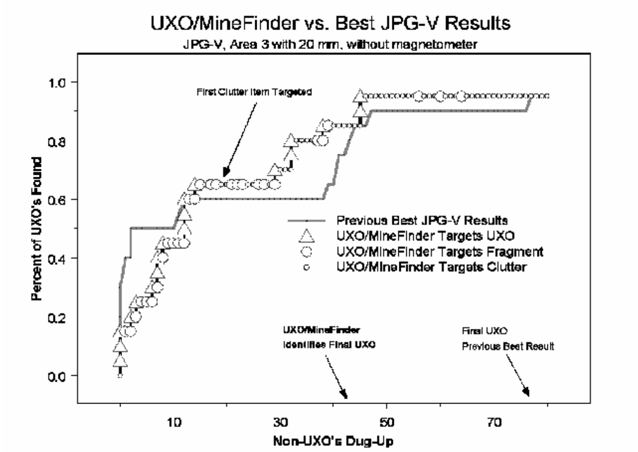

Fig. 3. Ranked Anomalies for JPG-V, Area 3. Comparison of LGP Based Rankings and Rankings by Previous Best Results for JPG-V, Area 3

Thus, if the order of digging were determined entirely by prioritization, and digging ceased when the last UXO was uncovered, the LGP based rankings would have required digging forty-five empty holes (that is, holes not containing a UXO) and the NRL rankings would have dug seventy-seven empty holes.

Digging up OE fragments is a secondary goal in UXO cleanup. Forty-five of the top sixty-four targets identified by our process contained OE fragments. In a field project, those fragments would be recovered in the process of digging up the UXO's. In fact, only nineteen truly empty anomalies were prioritized by LGP above the lowest priority UXO.

6. Future Work

In [17], we described an information theoretic optimal method to apply machine learning techniques to UXO discrimination across an entire site, even though no ground-truth is available at the start of the site clean-up. This technique permits *site-specific* discrimination that takes into account factors such as soil conditions and peculiarities of UXO distribution, munition type and depth on a particular site. Our next step will be to apply LGP in the site-specific manned outlined in [17].

7. Acknowledgments

The researchers would like to extend their thanks to Dr. Clinton W. Kelly III, manager of SAIC's Internal Research and Development division, who supported this research with management and technical analysis, and by providing corporate funding to get the concept off the drawing board and into productive research.

References

1. Report of the Defense Science Board Task Force on Unexploded Ordnance. Department of Defense. December (2003).
2. Banzhaf, W., Nordin, P. Keller, R. Francone, F.: Genetic Programming, an Introduction, Morgan Kaufman Publishers, Inc., San Francisco, CA (1998).
3. Francone, F. D., and Deschaine, L.M.: Extending the Boundaries of Design Optimization by Integrating Fast Optimization Techniques with Machine-Code-Based Linear Genetic Programming. In: Information Sciences Journal, Elsevier Press, In-press: Amsterdam, The Netherlands (2004).
4. Francone, F., Discipulus Owner's Manual. RML Technologies, Inc. (2002). Available at www.aimlearning.com.

5. Advanced UXO Detection/Discrimination Technology Demonstration, U.S. Army Jefferson Proving Ground, Madison, Indiana,. Technology Demonstration Plan, Washington, DC. Naval Explosive Ordnance Disposal Technology Division (NAVEOD-TECHDIV). (2000).
6. Cespedes, E.: Advanced UXO Detection/Discrimination Technology Demonstration—U.S. Army Jefferson Proving Ground, Madison, Indiana. US Army Corps of Engineers, Engineer Research and Development Center, ERDC/EL TR-01-20 (2001).
7. Geonics, Ltd., http://www.geonics.com/tdem.html (2004)
8. Geonics, Ltd., http://www.geonics.com/em61.html (2004).
9. J. Koza, J., Bennet, F., Andre, D., Keane, M.: Genetic Programming III, Morgan Kaufman, San Francisco, CA, (1999).
10. Nordin, P., Francone, F. Banzhaf, W.: Efficient Evolution of Machine Code for CISC Architectures Using Blocks And Homologous Crossover. In: Advances in Genetic Programming 3, MIT Press, Cambridge, MA (1998).
11. Nordin, P.: A Compiling Genetic Programming System that Directly Manipulates the Machine Code. In: K. Kinnear Jr. (Ed.), Advances in Genetic Programming, MIT Press, Cambridge, MA, 1994.
12. Nordin, P, Banzhaf, W.: Evolving Turing Complete Programs for a Register Machine with Self Modifying Code. In: Proceedings of Sixth International Conference of Genetic Algorithms, Morgan Kaufmann Publishers, Inc., 1995.
13. Fukunaga, A. Stechert, D. Mutz, A: A Genome Compiler for High Performance Genetic Programming. In: Proceedings of the Third Annual Genetic Programming Conference, Jet Propulsion Laboratories, California Institute of Technology Pasadena, CA, Morgan Kaufman Publishers (1998), pp. 86–94.
14. Francone, F., Nordin, P., Banzhaf, W.: Benchmarking the Generalization Capabilities of a Compiling Genetic Programming System using Sparse Data Sets. In: Koza et al. (Eds.), Proceedings of the First Annual Conference on Genetic Programming, Stanford, CA, (1996).
15. Deschaine, L. M., Hoover, R. A., Skibinski, J. N., Patel, J. J., Francone, F. D., Nordin, P. and Ades, M. J.: Using Machine Learning to Compliment and Extend the Accuracy of UXO Discrimination Beyond the Best Reported Results of the Jefferson Proving Ground Technology Demonstration. In: Proceedings of the Society for Modeling and Simulation International's Advanced Technology Simulation Conference, April 2002. San Diego, CA, USA (2002).
16. Hall, M.: Correlation-based Feature Selection for Machine Learning. PhD Dissertation. The University of Waikato, Hamilton New Zealand (1999).
17. Francone, F. D. and Deschaine L.M: Getting It Right at the Very Start - Building Project Models where Data is Expensive by Combining Human Expertise, Machine Learning and Information Theory. In: Proceedings of the Business and Industry Symposium, 2004 Advanced Simulation Technologies Conference (Arlington, Virginia, April 18 - 22, 2004), The Society for Modeling and Simulation International, 162-168