FINAL REPORT

UXO Discrimination Study
Pole Mountain Target and Maneuver Area, WY

ESTCP Project MR-201156

FEBRUARY 2012

Levi Kennedy
Signal Innovations Group, Inc.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 23-05-2012	**2. REPORT TYPE** Final Report	**3. DATES COVERED *(From - To)*** 10/2011-5/2012

4. TITLE AND SUBTITLE
UXO Discrimination Study Pole Mountain Target and Maneuver Area, WY: Final Report

5a. CONTRACT NUMBER
MR-201156

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Levi Kennedy(PI), Larry Carin, Todd Jobe, Xianyang Zhu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Signal Innovations Group, Inc.
4721 Emperor Blvd., Suite 330
Durham, NC 27703

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report details the application of the SIG statistical learning approach to UXO discrimination at Pole Mountain Target and Maneuver Area, Wyoming. This technology has been developed and validated under previous SERDP/ESTCP efforts by SIG and Duke University. Specific core technologies were used in this discrimination. One important subset of the classification technology tested in this demonstration was multi-task learning (MTL); where information from previous sites is incorporated into the current classification.

The multi-task classifier outperformed the single-task classifier. All performance objectives were achieved. All of the UXO were classified as targets. The number of false alarms was fewer than 30% of the total. An additional approach was tested, a generative model, that reduced the number of false alarms to 32. This represents about 2% of the total false alarms.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Levi Kennedy
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** SAR		22	**19b. TELEPHONE NUMBER *(include area code)*** 919-323-3456

Table of Contents

List of Acronyms

BENet	Bayesian Elastic Net
EMI	Electromagnetic Induction
MTL	Multi-task learning
NMAL	Non-myopic active learning
PNBC	Parametric Neighborhood-Based Classifier
ROC	Receiver Operating Characteristic
SBC	Sparse Bayesian Classifier
SIG	Signal Innovations Group, Inc.
SLO	San Luis Obispo Demonstration Site
STL	Single-Task Learning
TEMTADS	Time-domain Electromagnetic Multi-sensor Tower Array Detection System
UXO	Unexploded Ordnance

List of Figures

List of Tables

Executive Summary

Signal Innovations Group, Inc. (SIG) has previously demonstrated the effectiveness of site-specific statistical learning for smartly selecting labeled training data to maximize target discrimination. This report details the application of the SIG statistical learning approach to unexploded ordnance (UXO) discrimination for Pole Mountain Target and Maneuver Area (Pole Mountain), Wyoming. This technology has been developed and validated under previous SERDP/ESTCP efforts by SIG and Duke University. Specific core technologies were used in this discrimination. These technologies fall broadly into the four analysis categories: the sensor/target model, feature selection, classification, and active label selection. Feature selection was performed using the Bayesian Elastic Net which has the benefit of retaining correlated and informative features for classification.

Classification was performed using three approaches. Two were semi-supervised discrimination models. The first of these was the standard single-task learning approach that has been used on previous demonstrations. The second was a multi-task learning approach where information from previous sites is incorporated into the classifier. The third classification model, was not discriminative, but rather was a generative (i.e. target features were estimated directly rather than distinguishing target responses from clutter responses).

The objectives of the study were to maximize correct classification of UXO and non-UXO, specify a no-dig threshold, and minimize the number of anomalies that could not be analyzed. All objectives were met by each of the classification approaches. Multi-task learning required fewer training data for discrimination. Predictions based on the multi-task learning model also had fewer false alarms than the single-task model. The generative model, however, outperformed both of the discriminative approaches in terms of number of false alarms. With the generative approach, all the UXO were revealed with only 32 unnecessary digs.

The results of the demonstration highlight the need to use different modeling approaches at different sites. Future work will focus on using generative and discriminative approaches synergistically based on adaptive estimates of site difficulty.

1. Introduction

1.1. Background

Signal Innovations Group, Inc. (SIG) has previously demonstrated the effectiveness of site-specific statistical learning for smartly selecting labeled training data to maximize target discrimination. This report details the application of the SIG statistical learning approach to unexploded ordnance (UXO) discrimination for Pole Mountain Target and Maneuver Area (Pole Mountain), Wyoming. This technology has been developed and validated under previous SERDP efforts by SIG and Duke University.

Many current analysis approaches rely on expert scientists to make educated decisions at multiple points in the discrimination analysis process. This situation is not scalable, transferable, or cost effective. The SIG approach standardizes the options and creates a documented process flow that can be explicitly followed.

1.2. Objective of the Demonstration

The main technical objective of the Pole Mountain demonstration is to validate and substantially automate the SIG learning process using next-generation electromagnetic induction (EMI) sensor data for discriminating UXO. All elements of human interpretation and intuition are being incrementally constrained or removed from the process, resulting in an automated process, where all algorithm parameters and thresholds will either be determined by specified site parameters (i.e., expected or inferred munitions types) or by data-driven inferences (i.e., cross-validated operating threshold). In particular, SIG tested the viability of multi-task learning (MTL) for discriminating the site. MTL leverages labels from previous sites in a principled way.

2. Technology

SIG applied and matured each of the three key process phases that constitutes the SIG statistical learning approach to UXO discrimination - called the “SIG Isolate” process. The three phases of Isolate include: Phase I - feature extraction, Phase II – site learning, and Phase III – excavation. Each of the phases is described in detail below. Validation of Isolate entails meeting all of the discrimination performance objectives defined by the program office for each of the sites considered (see Table 1). The key technology in Phase II consists of a semi-supervised classifier that incorporates both labeled and unlabeled data from the site of interest to train the classifier. In Phase II, an active-learning framework adaptively requests samples from the current site with the goal of maximally reducing classifier prediction uncertainty. Additional site information is leveraged via MTL. SIG performed both MTL and single task learning (STL) at Pole Mountain.

2.1. Technology Description

The SIG Isolate process laid out in [5] can be summarized in the following ‘recipe’ (Figure 3):

- **Data Conditioning** - First, raw, unlabeled anomaly data are received.
- **Subspace Denoising** - The anomaly data is denoised to ensure robust performance for discriminating late time-gate features.
- **Feature Extraction** - A robust multi-anomaly dipole model is fitted to the data. The polarizability parameters from this fitting become the set from which features are drawn for classifier training. In addition to the time-domain polarizabilities, a set of 9 ‘rate’ features were calculated. These features were the calculated by fitting the time-domain polarizabilities of each axis to an exponential-decay model:

$$p_i = r_{1i} + r_{2i} e^{\frac{-t}{r_{3i}}}$$

where $i \in \{x, y, z\}$ is the current axis, p is the polarizability, t is time and $\{r_1, r_2, r_3\}$ are the fitted rate parameters. Though r_{1i} is unphysical, it is useful for adjusting for noise at late time gates and where odd responses would make the optimization difficult. The optimized values of the rate parameters were found using non-linear least squares.

- **Basis Selection -** A few of the many possible features are selected based on their physical interpretation as they relate to the anomaly, and, using these features, the most informative set of anomalies are selected via an information metric to begin classifier training.
- **Feature Set Augmentation -** The feature set is then augmented by adding early, mid and late time polarizabilities values.
- **Automated Features Selection -** For the now larger feature set, the most relevant set of features is selected using BENet.
- **Semi-supervised PNBC Training (STL or MTL) -** When the PNBC is trained only using data from the current site of interest, it is called Single Task Learning (STL). When the PNBC is trained for multiple sites simultaneously it is called MTL. For the Camp Butner demonstration only STL was used.
- **Non-myopic Active Learning -** Based on the estimates made with the PNBC classifier, a new set of anomalies will be selected for labeling using NMAL. The goal at this step is to maximize the information gain from new labels requested from the set of unlabeled anomalies. The process is repeated as the PNBC classifier adequately learns data manifold. The stopping criterion for the learning process is apparent when the remaining unlabeled data points have approximately equal information for improving the classifier. At which point, labeling any one anomaly is no better than any other.
- **Excavation Adapted Threshold Selection -** At this point, the highest probability UXO are selected for excavation and labels. The classier continues to be retrained when new labels are revealed. This process continues until the highest probability UXO items excavated are all found to be clutter at which point digging stops.

The process outlined above falls into 3 broad phases: Feature Extraction, Site Learning, and Excavation. Details on each phase are given in the next subsections. The SIG Isolate process is relatively linear save for two feedback steps. The first feedback is in training the semi-supervised classifier, where additional anomaly labels are requested until the classifier reaches sufficient stability. The second feedback is during the excavation of anomalies, where the classifier is retrained with additional labeled anomalies until either the UXO/clutter predictions become highly separable or until high probability anomalies are substantially revealed to be clutter upon excavation.

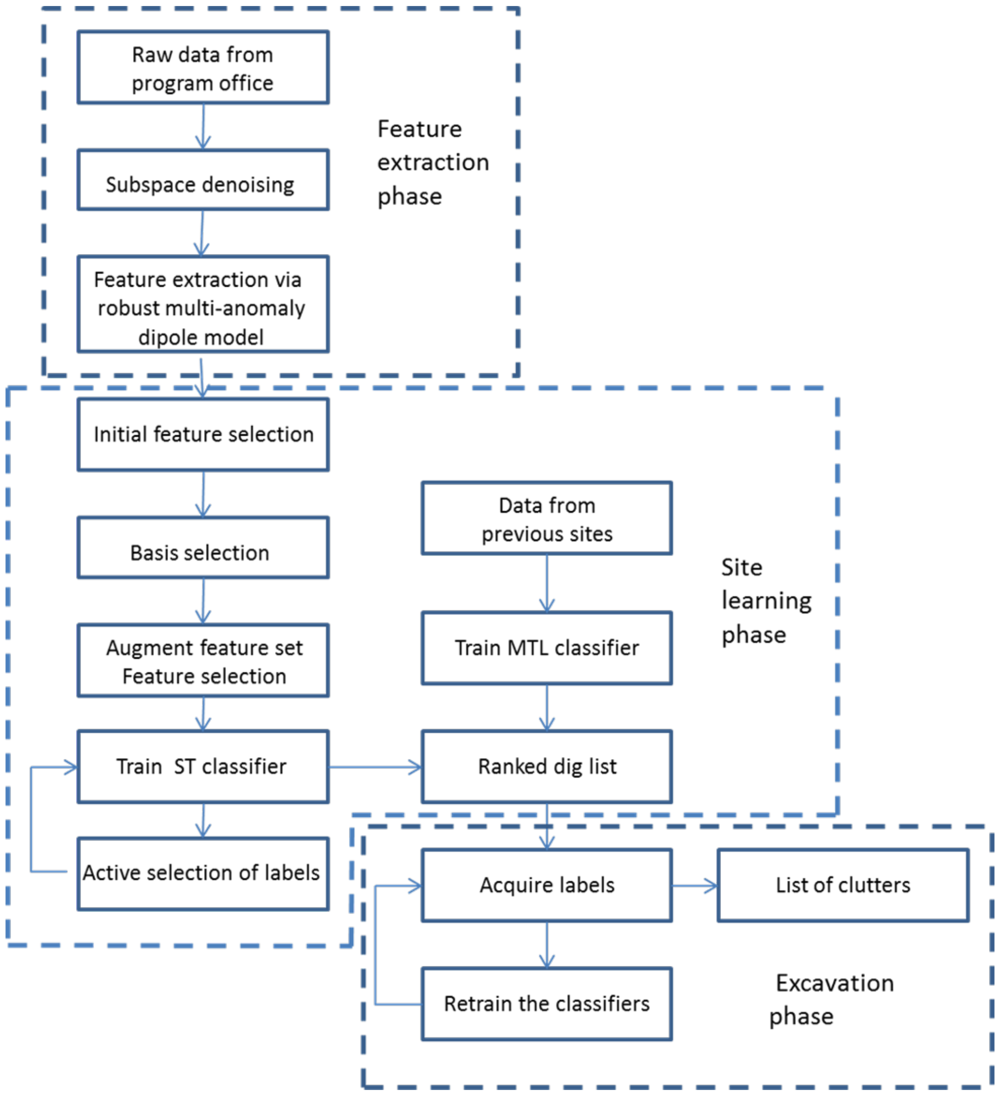

Figure 1. Flow diagram of the SIG Isolate process.

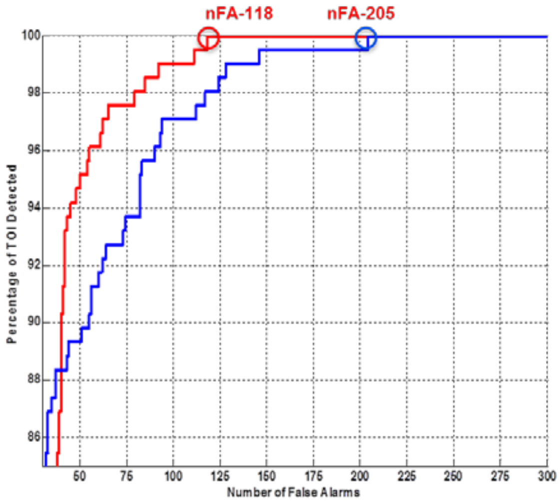

Figure 2: ROC curves for UXO classifier at SLO site with features selection using the BENet algorithm (red line) and without feature selection (blue line). The number of false alarms is lower for the classifier where feature selection was used.

2.2. Technology Development

SIG applied the Isolate discrimination process in the Camp Beale demonstration. This process involves the following key technologies including: Bayesian feature selection, semi-supervised classifier training, and non-myopic active selection of labeled data. These technologies are described briefly in the following subsections.

Feature Selection with BENet

Adaptive learning of a classifier *in situ* benefits from refining the appropriate set of extracted features for the targets under test. This occurs because of the 'curse of dimensionality' where the number of data points required to cover the breadth of a features space grows exponentially with the number of features considered. If the amount of training data does not sufficiently sample the feature space, then the learned classifier will lack statistical support and class estimate uncertainty is large. At the San Luis Obispo (SLO) demonstration site in particular, feature selection played a key role in classifier performance (Figure 1). Bayesian classification models perform feature selection by placing a sparseness prior on the inferred feature weights. The Bayesian elastic net (BENet) regression model used for feature selection employs a sparseness prior equivalent to a convex combination of $L1$-norm and $L2$-norm penalties in a least squares optimization formulation [1], [2]. The sparseness prior of the BENet model jointly infers the essential subset of relevant features, including correlated features, for a given classification task. Rather than encouraging the selection of a single feature in a set of correlated important features (like similar approaches such as RVM), the BENet model encourages the selection of all correlated important features. By performing sparse and grouped feature selection, the BENet algorithm provides a more robust approach to feature adaptability and the interpretation of important features, requiring fewer training data samples to achieve robust statistical support.

Semi-Supervised Classification

Semi-supervised learning is applicable to any sensing problem for which all of the labeled and unlabeled data are available at the same time, and therefore, particularly for the current demonstration study. In most practical applications (including the recent demonstration at Camp SLO), semi-supervised learning has been found to yield superior performance relative to widely applied supervised algorithms. Figure 2 depicts the advantage of a semi-supervised approach to classification over its supervised counterpart. A classifier trained purely on labeled data (depicted as red and green circles) is shown as a purple dashed line and generates classification errors. In contrast, a semi-supervised classifier trained on both labeled and unlabeled data will generate perfect classification (depicted by the blue line). Note that the context provided by the unlabeled data was crucial in improving the classification performance in this case, since the labeled data were not representative of the two class distributions. As the number of training samples increases, the supervised classifier should approximate the semi-supervised classifier. Semi-supervised formulation treats the dataset (labeled and unlabeled) as a set of connected nodes, where the affinity w_{ij} between any two feature vectors (nodes) $\boldsymbol{f_i}$ and $\boldsymbol{f}_j$ is defined in terms of a radial basis function [3]. Based on the above formulation, one can design a Markov transition matrix $\boldsymbol{A} = [a_{ij}]_{N\times N}$ that represents the probability of transitioning from node $\boldsymbol{f}_i$ to $\boldsymbol{f_j}$. Assuming $\mathcal{L} \subseteq \{1,2,\dots,N_L\}$ represents the set of labeled data indices, the likelihood functional can be written as

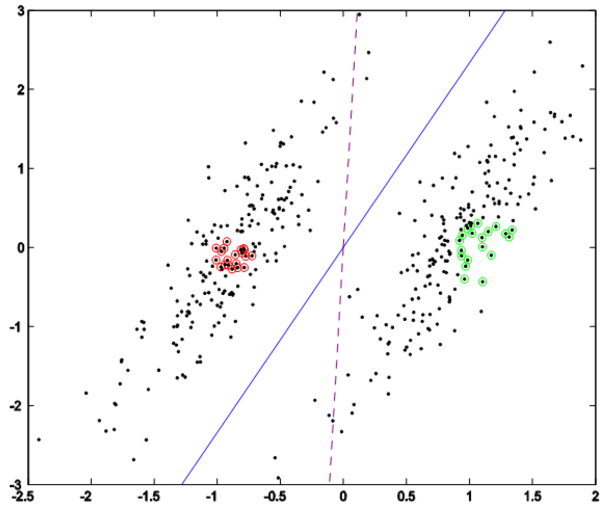

Figure 3: A comparison between supervised and semi-supervised classifiers for a 2. Labeled data from both classes (red and green circles) are shown, along with unlabeled data (black dots). The supervised classifier is trained on only the labeled data and the decision boundary is shown (dotted line). The semi-supervised classifier is trained on both the labeled and unlabeled data and the decision boundary (solid line) makes the two classes linearly separable.

$$(\{y_i, i \in \mathcal{L}\}|\mathcal{N}(\boldsymbol{f}_i), \boldsymbol{\theta}) = \prod_{i\in\mathcal{L}} p(y_i|\mathcal{N}(\boldsymbol{f}_i), \boldsymbol{\theta}) = \prod_{i\in\mathcal{L}} \sum_{j=1}^{N_i} a_{ij} p(y_i|\boldsymbol{f}_j, \boldsymbol{\theta})$$

where $\mathcal{N}(\boldsymbol{f})$ defines the neighborhood of $\boldsymbol{f}$. Estimation of classifier parameters $\boldsymbol{\theta}$ can be achieved by maximizing the log-likelihood via an Expectation-Maximization algorithm [4]. To enforce sparseness of $\boldsymbol{\theta}$ (enforcing most of the components of the parameter vector $\boldsymbol{\theta}$ to be zero), one may impose a zero-mean Gaussian prior on $\boldsymbol{\theta}$. A zero-mean Gaussian prior with appropriate variance can strongly bias the algorithm in choosing parameter weights that are most likely very small (close to zero). The algorithm we have used for this semi-supervised learning is termed a parameterized neighborhood-based classifier (PNBC).

Non-myopic Active Learning (NMAL)

Given that available training data labels at the beginning of a demonstration are not available and that excavations must be performed to reveal training data labels, one may ask in which order anomalies should be excavated to maximally improve the performance of the classifier algorithm. One useful criterion is to use the confidence on the estimated identity of the anomalies that are yet to be excavated. Specifically, one may ask which unlabeled anomaly label would be most informative to improve classifier performance if the associated label could be made available. It has been shown [5] that this question can be answered in a quantitative information-theoretic manner.

For active label selection, posterior distribution of the classifier is approximated as a Gaussian distribution centered on the maximum *a posteriori* estimate. The uncertainty of the classifier is quantified in terms of the posterior precision matrix. The objective of NMAL is to choose a feature vector for labeling that maximizes the mutual information (I) between the classifier $\boldsymbol{\theta}$ and the new data point to be labeled. The mutual information can be quantified as the expected decrease of the entropy of $\boldsymbol{\theta}$ after new sample $\boldsymbol{f}_{i*}$ and its label y_{i*} are observed.

$$I = \frac{1}{2}\log\frac{|H'|}{|H|} = \frac{1}{2}\log\{1 + p(y_{i*}|\boldsymbol{f}_{i*}, \boldsymbol{\theta}) \times [1 - p(y_{i*}|\boldsymbol{f}_{i*}, \boldsymbol{\theta})]\boldsymbol{f}_{i*}^T H^{-1}\boldsymbol{f}_{i*}\}$$

It is important to note that the mutual information I is large when $p(y_{i*}|\boldsymbol{x}_{i*}, \boldsymbol{\theta}) \approx 0.5$. Hence, the NMAL prefers to acquire labels on those unlabeled samples for which the current classifier is most confused or uncertain. In this fashion the classifier learns quickly by not excavating anomalies that reveal redundant information. The process continues as new labels are revealed until the expected information gain for the remaining anomalies is approximately uniformly low. At that point the classifier is adequately trained and target inference on the remaining unlabeled anomalies can be reliably performed. By invoking the principle of submodularity in the algorithm optimization, the approach has been adapted to allow for the selection of multiple simultaneous labels at one time, making the technique operationally practical.

Multi-Task Learning

SIG demonstrated a MTL classifier [4] for discrimination of TOIs, in which M parameterized classifiers, each associated with a demonstration site, are learned jointly while sharing a soft prior over the classifier parameters. Multi-task learning leverages information from past demonstrations, for example, data collected by Metalmapper from one site will be utilized in a principled way to design a classifier for subsequent sites that deploy Metalmapper. The MTL-based information sharing is crucial in training a classifier with a small amount of training data.

For example, suppose TEMTADS was deployed in Site 1 and the labels for all anomalies have already been revealed. If TEMTADS is deployed in Site 2, the MTL framework will utilize all labeled data from Site 1, along with Site 2 data to jointly train classifiers for both sites. This process does not pool data from multiple sites, but learns the classifiers for both sites in a manner that they influence each other. This process has already been shown to improve classification performance, while requiring fewer labeled samples from Site 2 [4]. SIG envisions that there would be considerable overlap between the sensors and munitions types found in the six demonstrations, and the MTL framework will allow the classification module for each sensor platform to leverage past information effectively to classify buried anomalies.

3. Performance Objectives

Performance objectives are summarized in Table 1. Each objective is described in a subsection below.

Table 1. Program Office Performance Objectives for Discrimination Analysis

Performance Objective	Metric	Data Required	Success Criteria
Analysis and Classification Objectives			
Maximize correct classification of targets of interest	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from the IDA	Approach correctly classifies all targets-of-interest
Maximize correct classification of non-UXO	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest
Specification of no-dig threshold	Probability of correct classification and number of false alarms at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 98% of anomalies on each sensor's detection list.

3.1. Maximize correct classification of targets of interest

A non-linear and a linear classifier were trained based on training labels requested from the program office. The objective was to predict all remaining UXO using the trained classifiers. This is measured by comparing the number of UXO captured from the dig list against the total number of UXO in the dataset. The necessary data are the dig lists and the scoring reports from the IDA. Some UXO were missed, and so the performance objective was evaluated in the context of how many additional digs would have been necessary to actually capture all the UXO.

3.2. Maximize correct classification of non-UXO

For both classifiers, a secondary objective is to capture all the UXO while keeping much of the clutter in the ground. Success was measured by keeping at least 70% of the clutter in the ground. Since, some UXO were left in the ground given the no-dig threshold, the number of false alarms was smaller than it should have been. This objective was re-evaluated in terms of how many false alarms would have been necessary were the digging thresholds set to capture all the UXO.

3.3. Specification of no-dig threshold

The objective was to give a reasoned operating point for splitting the dataset into anomalies that should be dug and those that should not be dug. The decision for this objective influenced the performance of the values in the first two objectives. The decision to stop digging was based on the separation between the posteriors predicted probabilities of the anomalies not used for training. The selected operating point based on this criterion left UXO in the ground.

3.4. Minimize the number of anomalies that cannot be analyzed

The objective was to have a minimal number of anomalies where the dipole inversion model gave poor results. This is a function of the data quality, something that was not controlled in this study, and a function of the efficacy of the inversion model. The decision to place anomalies in the 'can't analyze' category was based on the residual error of the least-squares model used for the dipole inversion. Anomalies with high residual error were removed. Success in this objective was defined as a creating effective parameterizations for >98% of the anomalies. This objective was achieved.

4. Site Description

All raw sensor data were provided to SIG directly. So there were no in-field components to the SIG discrimination.

5. Test Design

All raw sensor data were provided to SIG directly. So there were no in-field components to the SIG discrimination.

6. Data Analysis and Products

6.1. Training Steps

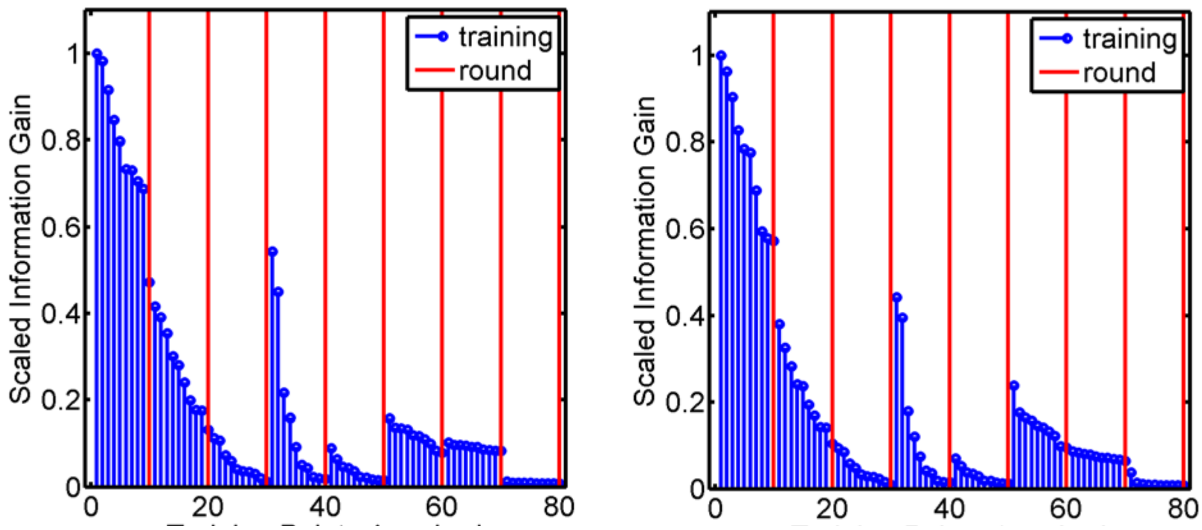

Figure 5. Fished information gain as a function of the training points acquired over 8 training rounds for STL (left) and MTL (right).

An initial basis of 20 vectors was selected maximizing Fisher Information gain. This initial selection was the same for both MTL and STL. From this initial sample, a set of relevant features was selected using BENet. Subsequently, non-linear PNBC classifiers using MTL and STL were trained on the original bases represented by these features. For the MTL classifier the additional tasks were MetalMapper data from different sites. These sites included Camp SLO, Camp Butner, and Camp Beale (the Beale Open site). Given the trained classifiers, a new set of 10 unlabeled anomalies were selected using batch NMAL. This is where the set of training observations for MTL and STL diverged. Each method selected slightly different anomalies for training via NMAL. Surprisingly, the selected labels were not drastically different. In the first round of training the MTL and STL share 8 out of the 10 labels requested. And as the training rounds progressed the methods would tend to request similar labels. These labels were not

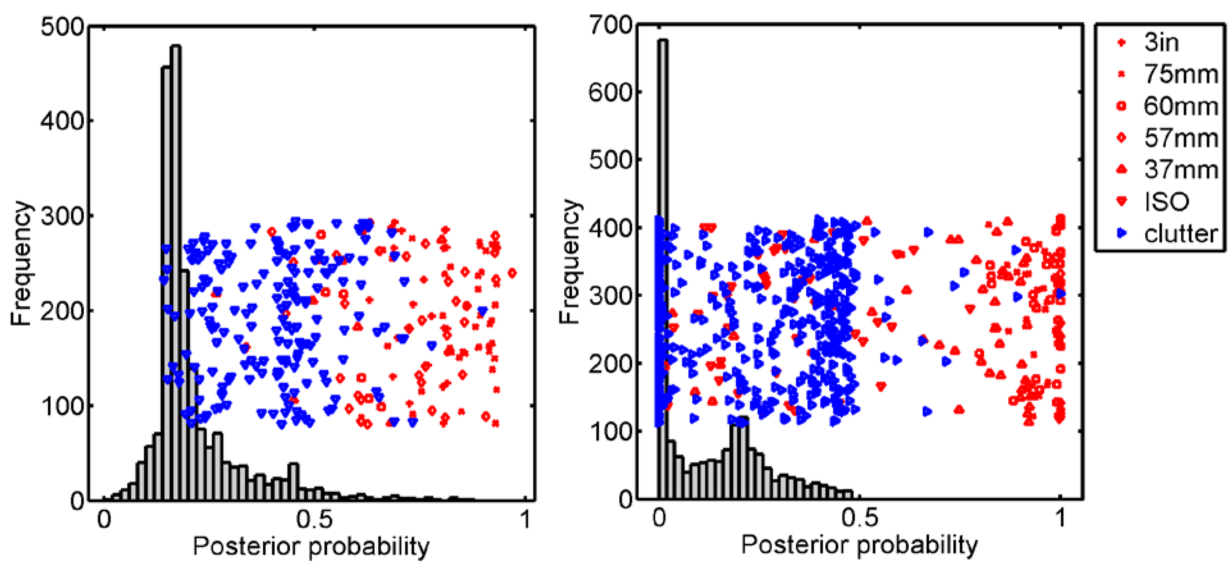

Figure 4. Histogram of predicted probability of being UXO at the end of NMAL training for STL(left) and MTL(right). Resubstitude prediction probabilities for the training data are also shown as filled shaped: clutter (blue), UXO (red).

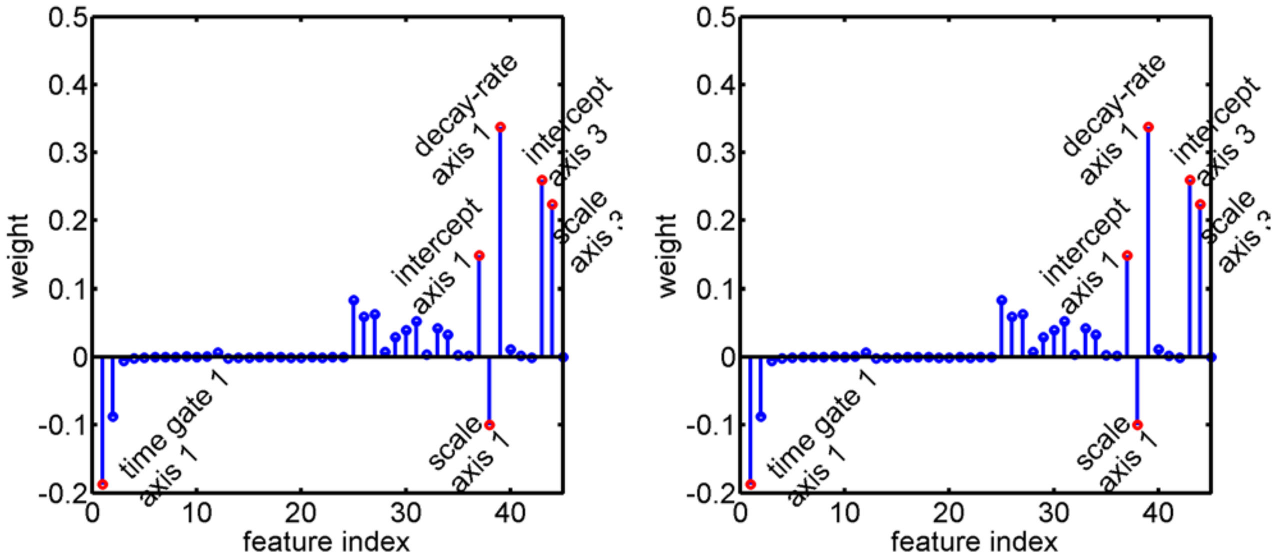

Figure 6. BENET feature weights at the end of training for STL (left) and MTL (right). Features used in the nonlinear classifier are highlighted in red and the feature names are given.

necessarily acquired in the same order or in the same training round, but at the end of training via NMAL both MTL and STL shared 98 out 100 training labels. The correlation in the labels selected suggests that the amount of new information contributed by the inclusion of the classifiers from the other sites (tasks) was minimal for this site. The decision to stop actively selecting labels was based on the decrease in Fisher Information gain and the separation between the predicted probabilities of UXO and clutter(Figure 4, Figure 5). For MTL there were no observations left whose predicted probability of being UXO was greater than 0.5.

For each round of label selection, feature selection was performed using BENet. These features converged to a fixed set by the end of training (Figure 6). Unlike classification with PNBC, feature selection with BENet was not multi-task. So, the MTL and STL classifications tended to use the same set of features. All the selected features were associated with either the magnitude or decay rate of the first and third polarizability axes. This is similar to the feature selection results from other sites (e.g. Camp SLO and Camp Butner).

Twelve rounds of training were performed for both MTL and STL. A dig list was then submitted for each of the two classifier algorithms. These dig lists missed many (> 40) quality control seeds. It was decided that instead of the program office giving SIG that many unrequested labels, another training round would be performed. Further, instead of focusing on acquiring informative labels based on NMAL the requested labels would be based on the posterior probability of being UXO. That is, the subsequent training rounds for MTL and STL requested labels for anomalies predicted to be UXO. Having received these labels, the models were retrained and additional labels were requested, again according to the probability of being UXO. This continued until a total of 15 training rounds for MTL and 12 training rounds for STL were completed. At this point the total number of training labels acquired for STL was 387 and 368 for MTL. There were no anomalies that were labeled ‘can’t analyze’ because of poor inversion results.

The predicted probability of being UXO at the end of training differed between MTL and STL even though the set of training data they acquired were similar (Figure 4). This occurred because the posterior predictions of MTL depend not only on the evidence presented by the Pole Mountain data, but also on the

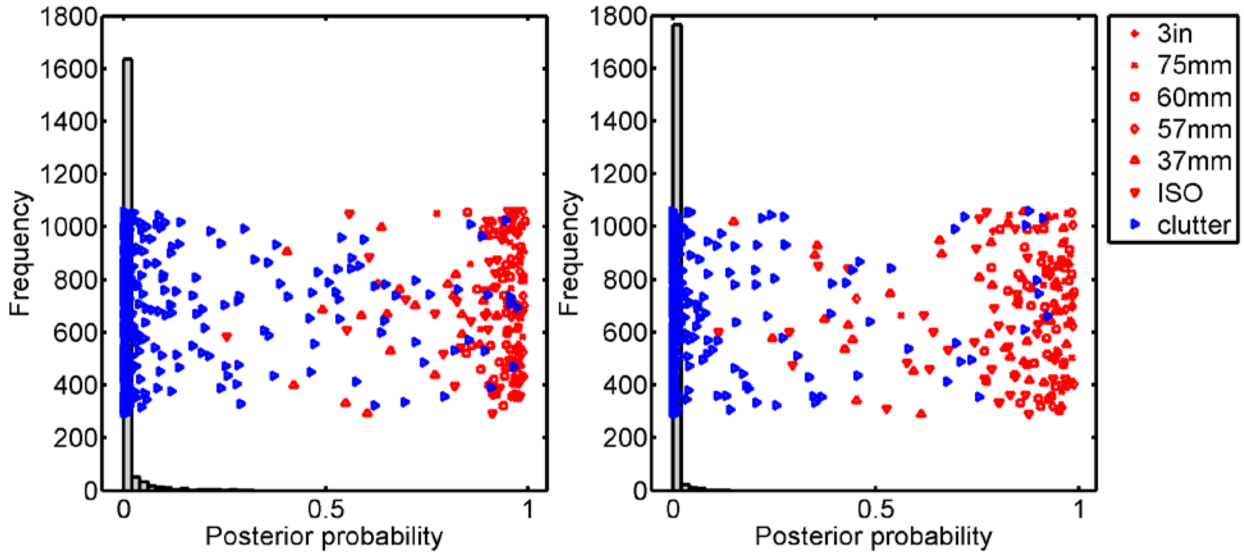

Figure 7. Histogram of predicted probability of being UXO at the end of digging for STL(left) and MTL(right). Resubstituted prediction probabilities for the training data are also shown as filled shaped: clutter (blue), UXO (red).

joint prior that includes the data from the other sites. A stage 2 dig lists was submitted for MTL and STL. These dig lists were the final lists submitted to the program office. The initial Stage 1 dig lists where many seeds were missed was counted in the number of dig lists submitted, but the labels were not sent from the program office for those lists. A total of 561 labels were requested for MTL . 582 labels were requested for STL.

7. Performance Assessment

No UXO were missed in the MTL and STL lists (Figure 8). Approximately 80% of the clutter was left in the ground for both methods. Fewer than 50 unnecessary digs were performed for both MTL and STL after the final UXO was dug.

MTL outperformed STL. Fewer training data were required for MTL before the final dig lists were submitted. MTL captured the last UXO with fewer unnecessary digs. Also, fewer labels were requested for MTL overall. These results are significant since the set of training data acquired by NMAL was basically identical for MTL and STL.

Even though MTL performed better than STL, both methods were ineffective at capturing the UXO efficiently. Discriminating the UXO from clutter at Pole Mountain should have been relatively 'easy' compared to other sites like Camp Beale because the inversion results were consistent. This begs the question, "Why did the SIG Isolate process require so many more digs to capture all the UXO?" The answer lies in the complexity of the tool used for discrimination and the difficulty of the site. Complex models require more representative training data to adequately describe, but are more effective at representing higher order data manifolds. The complex discrimination machinery used in the SIG Isolate process is most effective on sites that are difficult to discriminate. Alternative approaches that do not require much training data and focus on digging UXO rather avoiding clutter are better for easy sites. Once it was observed that this site was more separable, SIG employed a more appropriate data representation: a generative model.

The Generative Approach

Pole Mountain was a relatively easy site to classify. The original formulation of the SIG Isolate process was geared toward sites that are difficult to classify. Consequently, the MTL and STL classifiers, though all the UXO were captured, had too many unnecessary digs. To make the Isolate process more adaptable to easy sites like Pole Mountain, SIG extended it to include the option of a generative model. This generative approach was tested initially on the TEMTADS 2x2 data for Camp Beale with good results. A brief explanation distinguishing the discriminative and generative approaches is given below, along with the performance of the Pole Mountain classification using the generative model.

Classification approaches

There are two distinct approaches to classification: 1) the generative approach, and 2) the discriminative approach. The generative approach models the probability of being a target directly, without considering the distribution of clutter. The discriminative approach models the probability of being a target against the probability of being clutter. This is the technique used for classification by the MTL and STL performance assessments. One of the key benefits of using a generative approach is that digging can begin immediately from test pit data. In other words, no responses from clutter are necessary to train the model. The weakness of the generative approach is the possibility of missing hidden modes of UXO in the features space that would only be elucidated by exploring the clutter space as in the discriminative approach. There is no concept of NMAL in the generative approach because there is no classifier

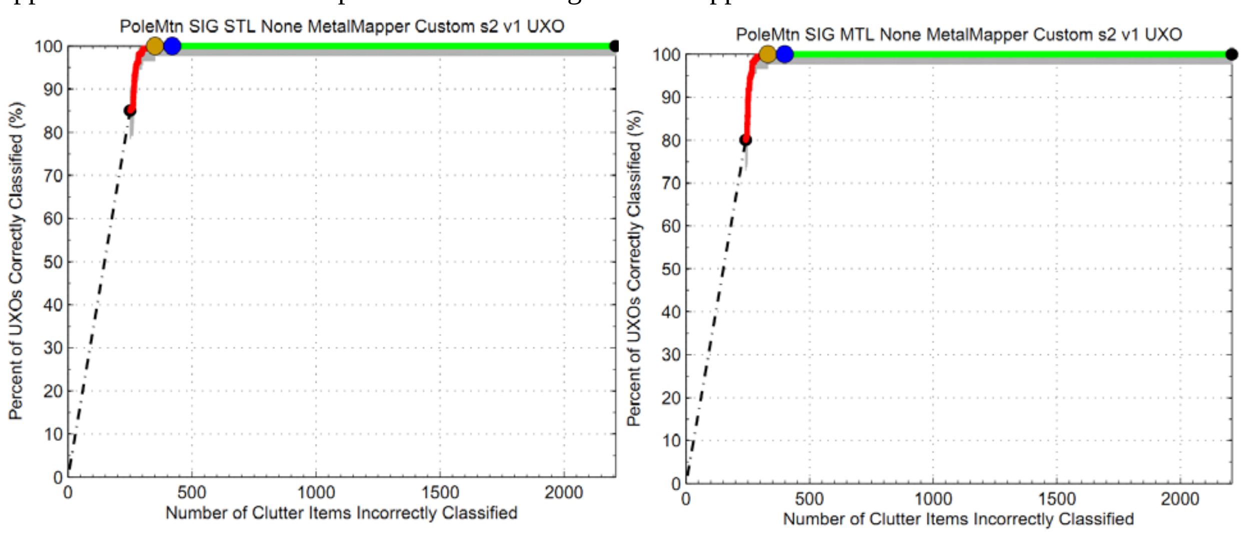

Figure 8. ROC curves for STL (left) and MTL (right) discrimination of Pole Mountain.

boundary *per se* to discriminate along. In performance assessment the ROC curve for generative approaches will, in general, be steeper initially than a discriminative approach. This is because training of generative models seeks only to find representative responses for UXO, whereas the discriminative approach seeks to find the boundary between UXO and clutter and explore areas of the feature space where little prior information is available.

The generative approach is completely dependent on the list of known UXO responses. So, if a hidden mode of UXO that does not exist in the known labels is present, then the number of false positives required to capture the last UXO will be greater than the discriminative approach. Given a generative model and a discriminative model with similar numbers of total clutter dug, the discriminative approach will tend to dig clutter during training and the generative approach will dig clutter toward the end of the 'dig phase'. This is obvious in the ROC curves for the discriminative results of MTL and STL where training accounts for half of the dug clutter (Figure 8).

Generative Model Training and Performance

There is really no distinction between 'training' labels and 'digging' labels in the generative approach. Since the UXO are being dug according to the highest probability of being UXO, the 'training' digs up likely UXO. So, all of the UXO were dug during the 'training' stage of the analysis. Labels were acquired over the course of 12 training rounds. The number of labels acquired in each training round varied from 24 in the first round to 5 in the final round, 12. A separate generative model was made for each UXO type. UXO were dug according to the probability of being UXO from greatest to least for each UXO type. When a given UXO type 'dug' 3 clutter in a row, then labels from that model were no longer acquired. The exception to this was if a UXO of that type was revealed later while digging a separate UXO type. For example if 3 clutter were dug in a row for the 37mm generative model, then no more digging would occur for the 37mm model. But, if another 37mm was revealed while digging according to the ISO model, then digging would begin again for the 37mm until 3 additional clutter were revealed. The number of clutter dug before a given model was stopped depended upon the number of training labels that were requested for a training round. In general, receiving fewer labels in a training round increased performance, but obviously increased the number of training rounds necessary to capture all of the UXO. Receiving more training labels per round increased the number of clutter that would be dug for a given UXO type before digging could stop for that type. 3 clutter labels per UXO type was chosen for Pole Mountain to keep the number of training rounds at or below 15. 12 training rounds were actually used.

Feature selection from the Pole Mountain data itself was not possible for the generative model. The reason for this is that the data acquired by the generative model are highly biased toward UXO with only a few clutter being dug. And, these clutter have responses that are very similar to UXO. Indeed, feature selection is only possible in a discriminative setting where the model distinguishes UXO from clutter directly. Instead, the features for the Pole Mountain generative model were selected based on previous sites. In this sense, the generative approach was similar to MTL in that information from other sites was leveraged in the classification. 5 features were used in the generative approach: the area of the object's transverse cross section, the object aspect ratio, the object symmetry, the magnitude of the 3rd axis polarizabilities, and the ratio between polarizabilities of the first axis' first and last time gates. These features were selected from a sparse Bayesian classifier [6] discrimination of the TEMATADS 2x2 data at Camp Beale. This was the only other site where the generative model had been applied and it was assumed that the features that were appropriate for the TEMTADS 2x2 data would also be appropriate for the Pole Mountain data. This turned out to be a good assumption and is probably extendable to other sites as well.

The performance of generative approach for classifying Pole Mountain was markedly better than the performance of the MTL and STL approaches. All the clutter was dug with only 32 unnecessary digs (Figure 9). The final dig list ended being named stage 2, but that was due to a simple typographical error in the stage 1 dig list. Having 32 unnecessary digs was not only much better than the MTL and STL approach, but was also better than most if not all of the other competitors.

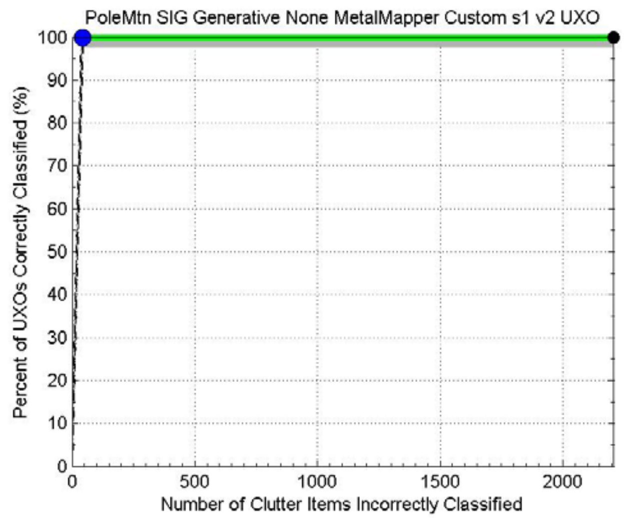

Figure 9. ROC Curve for the generative model. All UXO were dug during training.

Retrospective

The SIG Isolate process was improved greatly during the course of this analysis. A large benefit was shown by including previous site information through MTL. MTL incorporates this prior site information in a principled way so that future sites are not unduly influenced by previous sites. This technology will become key as the number of sites that have already been classified increases.

SIG has also demonstrated the benefit of using a generative approach at sites where the UXO are relatively easy to classify. The generative approach eliminates the need for training data to build a classifier, and begins digging UXO immediately. The degree to which the generative approach will minimize the number of unnecessary digs is dependent on how many UXO are dug before retraining occurs. For the performance assessment, all the UXO were dug with only 32 unnecessary digs, and the model was retraining after <25 labels were received. In a separate analysis, SIG used the generative approach where only a single label was acquired before the model was retrained. This represents a form of 'in-the-field' learning that could be incorporated into the digging protocol. This approach revealed all the UXO with only 3 clutter dug.

In future work SIG will develop a method for adaptively deciding whether a site will be difficult to classify or easy. Using this information the SIG Isolate process will be enhanced so that the generative approach and discriminative approach will applied along a gradient. There will be a continuous balance achieved between the generative predictions and discriminative predictions. When generative predictions are appropriate, likely UXO will be dug. Where discriminative predictions are appropriate, the classifier boundary will be refined via NMAL.

7.1. Maximize correct classification of targets of interest

The linear and non-linear classifications retained 163 and 164 UXO, respectively. This was the only performance object that was missed. It was missed due to a poorly chosen no-dig threshold. Were the stopping point moved to 625 false alarms, both methods would have met all of the performance objectives.

7.2. Maximize correct classification of non-UXO

If the dig-threshold were chosen correctly, then the reduction of false alarms would have been 75% for the linear classification and 85% for the non-linear classification. The no-dig threshold set too early, however. So, both classifications reduced the number of false alarms by 90%, but left UXO in the ground.

7.3. Specification of no-dig threshold

The operating point for the no-dig threshold was set at approximately 230 false alarms for both the linear and non-linear classifiers. The decision to stop digging was based on the separation between the posteriors predicted probabilities of the anomalies not used for training.

7.4. Minimize the number of anomalies that cannot be analyzed

98% of the anomalies had target parameters extracted effectively. 2% had large fit errors for the non-linear least squares model used for dipole inversion, were labeled "can't analyze", and marked for digging.

8. Cost Assessment

This section should provide sufficient cost information such that a professional involved in the field could reasonably estimate costs for implementation at a given site. In addition, this section should provide a discussion of the cost benefit of the technology. The following subsections with detailed discussions and examples should be provided.

8.1. Cost Model

The cost model is summarized in Table 2. The total cost per anomaly is $16.9. Each cost element is described in subsections below.

Table 2. Cost Model for the SIG Discrimination at Pole Mountain

Cost Element	Data Tracked During Demonstration	Estimated Costs
Feature Inversion	Unit: $ per anomaly • Time required • Personnel required • Number of sensors • Number of classifier techniques	10.4
Classifier Training/Testing	Unit: $ per anomaly • Time required • Personnel required • Number of sensors • Number of classifier techniques	3.9
Reporting	Unit: $ per anomaly • Time required • Personnel required • Number of sensors • Number of classifier techniques	2.6

Feature Inversion

Feature inversion includes any denoising and data preprocessing. The input data product here is the raw sensor data. The output data are the polarizabilities from the dipole model. Additional quality checks are performed at this stage. Costs would scale less than linearly with number of anomalies, because the time required for quality control is roughly the same regardless of the number of anomalies.

Classifier Training/Testing

Classifier training and testing encompasses all the data analysis required to move from anomaly polarizabilities to a final dig list. This includes requesting training data from the program office, feature selection, active learning, and quality assurance. Costs scale less than linearly with number of anomalies, because the percentage of training data required should decrease as the total number of anomalies increases.

Reporting

This in includes documentation of all feature inversion, classifier training/testing, and classifier performances. The cost should scale linearly with the sensors and classification techniques used.

8.2. Cost Drivers

The purpose of the SIG Isolate discrimination process is to decrease the cost per anomaly and to do so in a manner that scales well with production level discrimination. As the requirement for expert intervention and interpretation decreases, the scaling of the cost per anomaly should improve.

8.3. Cost Benefit

While the SIG Isolate process is not completely automated at this point, increasing automation drives the cost per anomaly toward becoming simply a function of computing time required and quality assurance checks. Since analyst time is the greatest cost in the discrimination process, automation provides excellent cost benefit for discrimination.

9. Implementation Issues

The software for the current SIG Isolate technology is based on MATLAB® and is not freely available. While the software is currently used by the experts who wrote the system, transitioning to minimally trained users is a goal of the software development. Future demonstrations will be used to mature this software.

10.References

[1] Y. Zhang, X. Liao, and L. Carin, “Detection of buried targets via active selection of labeled data: application to sensing subsurface UXO,” IEEE Transactions on Geoscience and Remote Sensing, vol. 42, pp. 2535–2543, November 2004.

[2] Y. Zhang, H. Y. L. Collins, C. Baum, and L. Carin, “Sensing of unexploded ordnance with magnetometer and induction data: Theory and signal processing,” IEEE Transactions on Geoscience and Remote Sensing, vol. 41, pp. 1005–1015, May 2003.

[3] Y. Zhang, X. Liao, and L. Carin, “Detection of buried targets via active selection of labeled data: application to sensing subsurface UXO,” IEEE Transactions on Geoscience and Remote Sensing, vol. 42, pp. 2535–2543, November 2004.

[4] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press, 2000.

[5] X. Liao and L. Carin, “Migratory Logistic Regression for Learning Concept Drift Between Two Data Sets With Application to UXO Sensing,” IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 12, December 2008.

[6] D. MacKay, “Information-based objective functions for active data selection,” Neural Computation, vol. 4, pp. 589–603, 1992.

11. Appendices

11.1. Appendix A: Points of Contact

POINT OF CONTACT Name	ORGANIZATION Name Address	Phone Fax E-mail	Role in Project
Levi Kennedy	Signal Innovations Group, Inc. 4721 Emperor Blvd., Suite 330 Durham, NC 27703	919-323-3456 919-287-2578 lkennedy@siginnovations.com	Principal Investigator
Lawrence Carin	Signal Innovations Group, Inc. 4721 Emperor Blvd., Suite 330 Durham, NC 27703	919 660-5270 919-323-4811 lcarin@ece.duke.edu	Project Management
Todd Jobe	Signal Innovations Group, Inc. 4721 Emperor Blvd., Suite 330 Durham, NC 27703	919-323-4811 919-287-2578 tjobe@siginnovations.com	Engineer
Xianyang Zhu	Signal Innovations Group, Inc. 4721 Emperor Blvd., Suite 330 Durham, NC 27703	919-323-4811 919-287-2578 xianyang@siginnovations.com	Engineer