Todo list

My Awesome Title

by

Arrielle C. Opotowsky

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Engineering Physics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2017

Date of final oral examination: 01/01/2100

The dissertation is approved by the following members of the Final Oral Committee:

Jane Doeverything, Professor, Electrical Engineering

John Dosomethings, Associate Professor, Electrical Engineering

Please insert your dedication here.

ACKNOWLEDGMENTS

It is customary for authors of academic books to include in their prefaces statements such as this: "I am indebted to ... for their invaluable help; however, any errors which remain are my sole responsibility." Occasionally an author will go further. Rather than say that if there are any mistakes then he is responsible for them, he will say that there will inevitably be some mistakes and he is responsible for them....

Although the shouldering of all responsibility is usually a social ritual, the admission that errors exist is not — it is often a sincere avowal of belief. But this appears to present a living and everyday example of a situation which philosophers have commonly dismissed as absurd; that it is sometimes rational to hold logically incompatible beliefs.

— David C. Makinson (1965)

Above is the famous "preface paradox," which illustrates how to use the `wbepi` environment for epigraphs at the beginning of chapters. You probably also want to thank the Academy.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

FIXME: basically a placeholder; do not believe

I did some research, read a bunch of papers, published a couple myself, (pick one):

1. ran some experiments and made some graphs,

2. proved some theorems

and now I have a job. I've assembled this document in the last couple of months so you will let me leave. Thanks!

1 INTRODUCTION

The realm of nuclear security involves many parallel efforts in nonproliferation (verification of treaty compliance, monitoring for smuggling, proper storage and transportation of nuclear materials), cyber security, minimizing stocks of weaponizable materials, disaster response training, and nuclear forensics. All of these efforts have been continually improving, but there was a gap regarding the ability of the United States (US) to coordinate and respond to a nuclear incident, especially with the technical portion of nuclear forensics: characterization and analysis. After all, the first textbook on the topic was published in 2005 . In 2006, the US Department of Homeland Security (DHS) founded the National Technical Nuclear Forensics Center (NTNFC) within the Domestic Nuclear Detection Office (DNDO). The mission of the NTNFC is to establish a robust nuclear forensics capability to attribute radioactive materials with demonstrable proof.

cite text

There are many fields that contribute to the nuclear forensics capability, such as radiochemical separations, material collection techniques, improving detector technology, material library development, and identifying forensic signatures. These needs vary based on whether the material being collected is post-detonation (i.e., bomb debris) or pre-detonation (i.e., spent nuclear fuel). In the pre-detonation realm, this project focuses on statistical methods to identify correlated material characteristics, which can lead to new forensic signatures.

1.1 Motivation

In the event of a nuclear incident, such as the retrieval of stolen nuclear material or the detonation of a dirty bomb, it is necessary to learn as much as possible about the source of the materials in a timely manner. In the case of non-detonated special nuclear material, knowing the reactor parameters that produced it can point investigators in the right direction in order to determine the chain of custody of the interdicted material. Determining these parameters (e.g., reactor type, cooling time, burnup) requires first characterizing and calculating certain isotopic ratios, chemical compounds, or trace elements. Both radiological methods (e.g., gamma spectroscopy) and ionization methods (e.g., mass spectrometry) measure these quantities. Although both measurement techniques have a multitude of techniques within them and thus varying strengths and weaknesses, the main tradeoff is between time/cost and amount of information gained.

The results of these analytic techniques are then compared against existing databases to obtain the desired reactor parameters. These databases are highly multidimensional, and furthermore, are rife with missing data entries and inconsistent uncertainties. Direct comparison between measurement results and a database therefore may not yield accurate results. Thus, computational techniques have been developed by nuclear engineers to calculate the parameters relevant to nuclear forensics analysis. Another approach with the uniqueness of requiring minimal domain knowledge is

Cite inverse stuff, possibly adjoint?

the use of statistical methods via machine learning algorithms to predict these characteristics or values. These algorithms can create a model using the database entries that enables "filling between the lines" of its entries. Additionally, having a machine-learned model may overcome the above challenges of multidimensionality, missing data, and irregular uncertainty.

1.1.1 Needs in Nuclear Forensics

The process of technical nuclear forensics includes the analysis and interpretation of nuclear material to determine its history, whether that be intercepted spent nuclear fuel (SNF), uranium ore concentrate (UOC), or the debris from an exploded nuclear device. After the technical portion is complete, intelligence data can be used to aid in material attribution; this is the overall goal of nuclear forensics.

After a nuclear incident, the material or debris is sampled and evaluated through many techniques that provide the following information: material structure, chemical and elemental measurements, and radioisotope measurements. These measurements or calculated ratios comprise the forensic signatures of the sample in question. These signatures can be analyzed with specific domain knowledge; for example, UOC will have trace elements depending on where it was mined from. They can also be analyzed against a forensics database, in the case of SNF.

Measurement needs and techniques vary vastly depending on the material, as does the type of signature. This study focuses on non-detonated materials,

specifically, SNF. It is important to determine if some intercepted material is from an undisclosed reactor or a commercial fuel cycle to attribute it to an entity or state. This is typically done by obtaining chemical and elemental signatures as well as isotopic ratios, and comparing these measurements to those in an existing forensics database of reference SNF.

The signatures for SNF correlate to several characteristics of quantities that can, in a best case scenario, point to the exact reactor from which the fuel was intercepted. The reactor parameters of interest are the reactor type, fuel type and enrichment at beginning of irradiation, cooling time, burnup.

The current and future work of this study are designed based on two primary needs to bolster the US nuclear forensics capability: forensics databases are imperfect, and our best measurement techniques are not always feasible in an emergency scenario. It is proposed that using a machine-learned model may be able to combat these issues, discussed in Section .

Forensics databases are kept by individual countries, and a given database will have widely varying uncertainty depending where and on what instrument the material was measured. Additionally, some fields have missing data. This presents issues with matching and characterizing SNF based on interpolating between entries.

A lofty goal for the forensics community would be to develop methods that provide instantaneous information, reliable enough to guide an investigation (e.g., within 24 hours). In the case of SNF, it takes weeks in a lab to

need citation, prob ORNL paper

more deets here? prob better later

ref

more deets and cite

measure isotopes via advanced gamma spectroscopy and mass spectrometry equipment. Thus, fast measurements to provide isotopic ratios to calculate the above-mentioned fuel parameters of interest would provide this via some form of a handheld detector that measures gamma spectra. However, while this nondestructive analysis is rapid, it is also difficult to evaluate because of the presence of overlapping peaks. Thus, gamma spectra give less information at a higher uncertainty than the near-perfect results of some destructive mass spectrometry techniques, like TIMS. Additionally, within gamma spectroscopy techniques (e.g., field vs. lab detectors), uncertainties can vary significantly because of the detector reponse, environment, storage, electronics, etc.

find MS technique and citation

1.1.2 Contribution of Statistical Methods

While different machine learning algorithms and parameters will be investigated, it is first important to determine if statistical methods can overcome the inherent database deficiencies. Thus, this paper focuses on probing the amount of information required to obtain realistic results. This can be best understood as the analgous real-world scenario. Although mass spectroscopy techniques provide extremely accurate isotopic information, they are time-consuming and more expensive. And although gamma spectroscopy can give extremely fast results cheaply, it only measures certain radiological signals and is influenced by many environmental factors, storage, and self-attenuation.

fix up and summary

In the simulation and machine learning paradigm, we need to determine what exactly is needed to train a machine-learned model. Of interest to an entity trying to create a weapon is partially irradiated fuel if they have plutonium separations capabilities or any radioactive substance in the case of a dirty bomb. Addressing the former, we used a set of simulations of spent nuclear fuel at different burnups and cooling times.

assumes previous section on fuel cycle sim

Can the algorithm overcome the deficiencies of gamma detection and still provide useful results? Or does it need more information, e.g., exact isotopics? Thus, ultimately, the goal is to answer the question *How does the ability to determine forensic-relevant spent nuclear fuel attributes degrade as less information is available?*.

But first, we must establish some baseline expectations of reactor parameter prediction and algorithms to use. This work is based off previous work on the subject (cite Dayman) regarding machine learning performace with respect to information reduction, and expands upon it by also evaluating a more advanced machine learning algorithm: support vector regression. Below is a more in depth discussion of nuclear forensics and how machine learning can contribute to this research area. After that, an experimental design is outlined. Lastly, the results are presented and discussed.

Given imperfect data with varying amounts of uncertainty as well as the required comparison to highly multidimensional databases with missing entries, many have begun considering computational approaches to nuclear forensics problems, such as the INDEPTH

cite and discuss

Another approach utilizes artificial intelligence to solve nuclear forensics problems, such as implementing searching algorithms for database comparison and machine learning for determining spent fuel characteristics . A variety of statistical and machine learning tools have been used to characterize spent fuel by predicting categories or labels (reactor type, fuel type) as well as predicting values (burnup, initial enrichment, or cooling time) The former uses classification algorithms and the latter uses regression algorithms. Many algorithms can be applied to both cases.

cite hungarian guy, viz guy, dayman paper

A typical (supervised) machine learning workflow would take a set of training data with labels or values inserted into some statistical learner, calculate some objective, minimize or maximize that objective, and provide some model based on that output. Then a test set (with known values) is provided to the model so that its performance can be evaluated and finalized. After model finalization, a user can provide a single instance and a value can be predicted from that.

insert ML schematic

To obtain reliable models, one must 1. choose/create a training set carefully and 2. study the impact of various algorithm parameters on the error. Many algorithms are developed on an assumption that the training set will be independent and identially distributed (i.i.d.). [Aside: there are ways to handle skewed data sets] This is important so that the model does not overvalue or overfit a certain area in the training space. Additionally, algorithm performance (or error) can be optimized with respect to training set size, number of features, or algorithm parameters (regularization terms,

etc). These are known as diagnostic plots. When plotting the training and testing error with respect to the number of instances, this is known as a learning curve. When plotting these errors with respect to the number or features or algorithm parameters, this is known as a validation curve.

Algorithm choice is usually based on what is being predicted and intuition regarding strengths and weaknesses. For the sake of comparison (i.e. weak validation), some machine learning approaches here are based on previous workwhile also extending to a more complex model via an algorithm that is known to handle highly dimensional data sets well. Thus, this paper investigates three regression algorithms: nearest neighbor, ridge, and support vectors.

insert example diagnostic plot?

cite dayaman

1.2 Methodology

Talk about workflow here. Might need separate tex file. Prob should keep pretty simple since everything is discussed later in the demonstration part of the prelim.

1.3 Goals

The main purpose of this work is to evaluate the utility of statistical methods as an approach to determine nuclear forensics-relevant quantities as less information is available. Machine learning algorithms will be used to train

models to provide these values (e.g., reactor type, time since irradiation, burnup) from the available information. The training data will be simulated using the ORIGEN tool , which will provide an array of nuclide concentrations as the features (X) and the value of interest (y) is provided from the simulation inputs. Information reduction will be carried out using computationally generated gamma spectra; the radionuclide concentrations from the simulations can be converted into gamma energies, which then undergo a detector response calculation to represent real gamma spectra as closely as possible. Machine learning best practices will be used to evaluate the performance of the chosen algorithms, and established statistical methods will be used to provide an interval of confidence in the model predictions .

cite

correct lang?

The necessary background is covered in Chapter 2. First, an introduction to the broader field of nuclear forensics is in Section 2.1 to place this work in the context of the technical mission areas. After that, a short discussion of the field of machine learning, the algorithms used, and validation methods are in Section 2.2. Section 2.3 includes information about the codes used to generate the training data, via fuel cycle simulation, detector response function, and isotope identification of gamma spectra. Lastly, a review of statistical methods being used in studies of forensics analysis is covered next in Section 2.4.

After the existing work is discussed and the gap that this work will fill is identified, the experimental components are introduced next in Chapter 4. A demonstration of the methodology is presented next in Chapter 3.

Following these two chapters, Chapter xx describes the thesis research proposal and corresponding hypotheses. Finally, future directions and alternative directions are identified in Chapter xx.

2 BACKGROUND AND LITERATURE REVIEW

This chapter provides a background and literature review of the necessary components for this project. Section 2.1 outlines the broader field of technical nuclear forensics, with a focus on the area that motivates this project. Section 2.2 introduces the field of machine learning for an uninitiated audience, covers the relevant algorithms, and presents the methods field practitioners use for validation. Next, Section 2.3 covers the computational methods used to generate the training data for the machine learning input. Finally, the marriage of Sections 2.1, 2.2, and 2.3 is presented in 2.4, which is a review of previous work applying statistical methods to the nuclear forensics analysis of pre-detonated nuclear materials.

2.1 Nuclear Forensics

I'm a section on the field of nuclear forensics.

2.1.1 Pre-Detonation Nuclear Forensics

2.1.2 Post-Detonation Nuclear Forensics

2.2 Machine Learning

Machine learning is a sub-field of artificial intelligence (AI) within the broad category of computer science. The goal of AI is to create computer systems

that respond to their environment according to some set of criteria or goal. For example, self-driving vehicles have computers on board that learn to avoid curbs and humans. It is common knowledge that the use of AI has been expanding at a rapid rate in recent years. News stories of major tech companies' AI advancements are frequent and news articles abound with data on which jobs will be replaced with AI in the near future.

While its use has been increasing in the commercial sector, there is also much anecdotal evidence to support the existence of a rapid increase of AI use in academic research across many disciplines beyond robotics. AI systems have been used in detection (e.g., fraud or spam), medical diagnostics, user analysis (e.g., Netflix ratings), and a host of scientific disciplines that have increasing amounts of multivariate data.

Much of the recent advances to the field of AI have occured in the statistical realm, which forgoes domain knowledge in favor of large data sets. Thus, machine learning and statistical learning has become somewhat of a separate field. Machine learning research focuses on the underlying algorithms using mathematical optimization, methods for pattern recognition, and computational statistics. As an application, however, this study is not concerned with computational time, but rather the ability to correctly predict values and categories relevant to the nuclear forensics mission. This restricts the relevancy of the algorithms to the underlying theory and its impact on the resulting model's accuracy.

could explain more, cite changing science of ML

Machine learning algorithms can be separated into two main categories:

unsupervised and supervised learning. The former groups or interprets a set of input data, predicting patterns or structures. The latter includes both the input and output data, enabling the trained model to predict future outputs. Broadly speaking, the unsupervised learning algorithms are designed for clustering data sets or dimensionality reduction (i.e., determining some subset or linear combination of features most relevant to the input data) of data sets. Supervised learning algorithms predict both discrete and continuous values via classification and regression, respectively. Some algorithms can perform both classification and regression, and neural networks can even be modified to perform either supervised or unsupervised learning. Additionally, various algorithms can be strung together, which is referred to as ensemble methods. One common way of doing this is performing deimensionality reduction prior to supervised learning.

This study performs both classification and regression tasks using supervised learning algorithms. Differences among the underlying mathematics of the algorithms do impact the performance of a machine-learned model. Therefore the algorithms used in this study will be discussed in 2.2.1. Additionally, evaluating algorithm performance and robustly comparing them is important, and this discussion is in 2.2.2.

check that this common workflow qualifies as ensemble

phrase better

2.2.1 Classification and Regression Algorithms

For relevant nuclear forensics predictions, both classification and regression algorithms must be used. For example, one may want to predict the reactor

type based on some measurements (referred to as features) of spent fuel of an unknown source, and this would require a classification algorithm. Or perhaps the input fuel composition is relevant to an investigation on weapons intent, so a regression algorithm would be used to train a model based on some set of features. Since algorithm formulation impacts the resulting performance, they are discussed in detail below.

should add useful vocab: training data X and y, instances, features, etc

2.2.1.1 Linear Models

One of the simplest and most obvious methods of prediction is a linear model using a least-squares fit.

not sure about this organization

2.2.1.2 Nearest Neighbor Methods

Nearest neighrbor is

$$\hat{Y}(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \tag{2.1}$$

intro, alg math, explain regularization

Nearest neighbor regression calculates a value based on the instance that is closest to it. The metrics for distance differ, but in this study, Euclidian

explain distance metrics

distance was used. There is no learning in this regression, per se; the training set populates a space and the testing set is compared directly to that. [2]

2.2.1.3 Support Vector Machines

Support vector regression (SVR) is an extension of the popular classification algorithm, support vector machine (SVM). This algorithm was chosen because of its ability to handle highly dimensional data well, which in this study is approximately 300 features.

add schematics from pres...or diff pics with the math?

SVM classifies two classes by determining an optimal hyperplane, given by wx+b, between them. As seen in Figure ?, the algorithm evaluates the quality of the line that separates two classes by maximizing the width of the margin given the contraints surrounding the line. Some problems are not linearly separable, and thus a penalty term is introduced to allow for misclassifications. As shown in Figure ?, the algorithm then simultaneously minimizes the misclassifications while maximizing the margin.

This can be extended easily to multidimensional analysis via what is called the *kernel trick*. First, using a nonlinear kernel function maps the data into higher dimensional feature space. Then the algorithm can find a linear separation in this space, as shown in Figure ?. Further, this can be upgraded from classification to SVR by doing similar math but instead minimizing the margin, as shown in Figure ?.

The kernel chosen for this study is the Gaussian radial basis function,

shown below. This has two tuneable parameters, gamma and C. Gamma influences the width of influence of individual training instances, and strongly affects the fitting of the model. Low values correspond to underfitting because the instances have too large of a radius (low influence) and high values correspond to overfitting because the instances have a small radius (high influence).

The C parameter also affects the fitting of the model by allowing more or less support vectors, corresponding to more or less misclassification. A lower C smooths the surface of the model by allowing more misclassifications, whereas a higher C classifies more training examples by allowing fewer misclassifications. Too low and too high of a C can cause under- and overfitting, respoectively.

Since there is a tradeoff of fitting strength provided by both parameters, it is common to run the algorithm on a logarithmic grid from 10'-3 to 10'3 for each parameter. If plotted on a heatmap of accuracies given gamma and C, there will be a diagonal of ideal combinations that emerges. The lowest of these is usually chosen.

2.2.1.4 Artificial Neural Networks

dlllllll

2.2.2 Comparing Models and Model Selection

bias-variance tradeoff???

2.3 Computational Methods

I'm a section on the computational methods!

2.3.1 Fuel Cycle Simulation

Short blurb on FC sim (capabilities enable other materials to be studied beyond SNF)

2.3.2 Data Modification

Perfect data -> real world data

2.3.2.1 Detector Response Functions

GADRAS?

Discuss methods, and potential options. Am I implementing for prelim?

2.3.2.2 Isotope Identification from Gamma Spectra

Same as above, but prob won't implement for prelim.

2.4 Applications of Statistical Methods to Nuclear Forensics Analysis

I'm the lit review section on previous applied ML work in the NF field.

2.4.1 Special Nuclear Materials Studied

The review on nf for the whole fuel cycle is useful here, perhaps. This is also important when I discuss my risk management section later.

2.4.2 Statistical Methods Employed

Very short details from lit review outline and success rates should be discussed here

3 DEMONSTRATION

Fill me in, yo.

3.1 Experimental Components

This work begins by simulating the training and test sets described in ref (cite Dayman). As with the previous work, this will be done using SCALE 6.2 . Specifically, the ARP module of the activation and depletion code ORIGEN was used.

The parameters of the training set are defined as follows. A smaller burnup than is typical for spent fuel from a commercial reactor is used in the previous work likely because stolen fuel pins for weapons use would not likely be at the end of their lifetime, as the plutonium of interest has decreased by then. A truly i.i.d. training set would go beyond this, but this is purely for demonstration with a single use case in mind.

The previous work also used an external test set, designed to have values in between the trained values of burnup. This is implemented in this study but it is expected that cross-validation will better indicate the model performance. More specifically, using k-fold cross-validation is a common method to use in the application of machine learning to create more confidence in the resulting model.

detector response function and iso ID

cite

add deets and citations

show table of training set space

show table of test set space

either explain or cite cross validation

here's some awesome success with only isotopes

3.1.1 Information Reduction

This study evaluates the impact of randomly introduced error of varying amounts on the ability of the algorithms to correctly predict the burnup. Thus, first investigated are the three algorithms with no error introduced, and next with the random errors applied uniformly to each nuclide vector. Since error in a nuclide vector is not random, in fact it is systematic and dependent on a number of known sources of uncertainty, the next study will introduce error by limiting the nuclides to only those that can be measured with a gamma spectrometer (future work).

3.2 Expanding the Training Set

BeepBop

4 METHODOLOGY

Some introduction here

4.1 Experiment

4.2 Experimental Design

Intro here?

4.2.1 Algorithm Evaluation

Additionally, machine learning algorithms are heavily dependent on the inputs and parameters given to them, such as training set sizes, learning rates, regularization, etc. To evaluate the performance or tweak the model from an algorithm, diagnostic plots will be used. Learning and validation curves will indicate how the models are performing, initially both with respect to the testing error and the cross validation error. As previously mentioned, these two errors are to be compared to the training error to understand the prediction and generalization strength with respect to training set size and the algorithm parameters governing model complexity.

The learning curves were obtained as follows. For a given (randomly chosen) training set size between 15 and 100% of the total data set, several training and prediction rounds were performed. The repetition for obtaining

the testing error is the same value as the k in k-fold cross validation. The testing error scenario averages the values of the obtained errors whereas the k-fold cross-validation performs this automatically. The validation curves were obtained as follows. For a given parameter, the value of the parameter is varied and k training and prediction phases are completed, and their errors averaged. Again, for k-fold cross-validation, these errors are already averaged. The learning curves help determine if we are over- or under-training. The validation curves help determine the optimal way to be robust to over- and under-fitting.

4.2.2 Comparing Algorithms

Maybe can delete this if I already discuss it in the above section. Options for comparison of algorithms: Comparing classification of 2 classes on same ROC plot with multiple ML systems, Scatter plots, Pairwise t-tests.

4.3 Posting here for now

1. Learning curve 2. Validation curve 3. Random error curve

Want to show the difference between test/train error plots and CV/train error plots. Determine some argument that perfers the latter. For the above categories in the validation section, can show plots directly next to each other (for 1, 2, 3) to hopefully show that cross validation provides better generalizability.

4.4 Validation

As previously mentioned in the demonstration section, the validation was carried out as a comparison of accuracies of the predicted classes (reactor type,) and values (enrichment amount) to the classes and values in the predetermined test data set (discussed in).

However, because it is difficult to ensure consistently representative testing data, the accuracy of a learned model should not depend on only one testing set . The learned model's accuracy can be better evaluated as discussed in by using cross validation . Other additional evaluation methods will be discussed here as well.

In addition to evaluating a single learned model, it may be beneficial to compare models. As discussed in , there are three methods that will be used: comparison of Receiver Operating Characteristics (ROC) curves, scatter plots, and pairwise t-tests.

Discuss understanding confidence intervals in predictions.

add label

other classification tasks?

or burnup?

add label when this section is done

general knowledge or citation?

REFERENCES

[1] Dayman, Kenneth, and Steven Biegalski. 2013. Feasibility of fuel cycle characterization using multiple nuclide signatures. *Journal of Radioanalytical and Nuclear Chemistry* 296:195–201.

[2] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2001. *The elements of statistical learning*. Springer Series in Statistics, New York, NY, USA: Springer New York Inc.