Evaluating Statistical Method Training Sets for Nuclear Forensics Analysis

Arrielle C. Opotowsky,* Charles F. Weber,† and Paul P.H. Wilson,*

*Computational Nuclear Engineering Research Group, University of Wisconsin–Madison, Madison, WI, opotowsky@wisc.edu, paul.wilson@wisc.edu

†Nuclear Security Modeling Group, Oak Ridge National Laboratory, Oak Ridge, TN, webercf@ornl.gov

ABSTRACT

One technical nuclear forensics area warranting research is the provenance of non-detonated special nuclear material. Studied here is spent nuclear fuel, which is applicable in a scenario involving the unlawful use of commercial byproducts from nuclear power reactors. By measuring known forensics signatures, one can ascertain the reactor parameters (e.g., reactor type, time since irradiation, burnup) that produced the material and therefore locate the source of the material. However, reducing the required domain knowledge to complete this task as well as increasing the speed of material attribution are important for an expedient nuclear forensics investigation. Additionally, reactor history databases used in nuclear forensics have a number of challenges: missing entries, inconsistent uncertainties, and a large number of dimensions. This work is proposing the use of statistical methods to determine reactor operation parameters instead of empirical relationships. This has the potential to determine correlations that are not used currently (most likely as new reactor technologies are deployed since SNF is so well studied). Thus, a proof-of-principle evaluation on the utility of statistical methods as an approach to determine nuclear forensics-relevant quantities is warranted.

This work emplys three machine learning algorithms: nearest neighbor, linear, and support vector regression. They are used to train statistical models to predict fuel uranium enrichment, burnup, and cooling time given some nuclide vector. The training dataset is simulated using the Oak Ridge Isotope Generation code, which provides an array of nuclide concentrations as features. The three parameters of interest are provided by the simulation inputs. The models are assessed according to their error of prediction. Although the nuclide vectors in this work are more populated than any assay in the real world could be, it is useful to evaluate the feasibility of this workflow. Before pursuing more physically realistic scenarios, the training set space needs to be studied. For example, certain parts may be more difficulty to predict.

The three algorithms (nearest neighbor, linear, and support vector regression) were chosen for their increasing levels of model complexity. After optimizing the respective algorithm parameters, the prediction error will be evaluated with respect to its location in the training set space. Additionally, various representations of prediction error will be discussed to illuminate the robustness of this proposed method to the method of error calculation.

ADD RESULTS

INTRODUCTION

In the event of a nuclear incident, such as the retrieval of stolen nuclear material or the detonation of a radiological dispersion device, 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., 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 spectroscopy) 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 requiring minimal domain knowledge is the use of statistical methods via machine learning (ML) algorithms. These algorithms can create a non-empirical model using the database entries that enables "filling between the lines" of its entries. Additionally, having an ML model may overcome the above challenges of multidimensionality, missing data, and irregular uncertainty.

While different ML approaches and algorithm parameters will be investigated, it is first important to determine if statistical methods can overcome the inherent database challenges/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. Mass spectroscopy techniques provide extremely accurate isotopic information, but 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.

In the simulation and ML paradigm, we need to determine what exactly is needed to train an ML 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, this work uses a set of simulations of spent nuclear fuel at different burnups and cooling times. The main goal is to answer the question *How does the ability to determine forensic-relevant spent nuclear fuel attributes degrade as less information is available?*.

Nuclear Forensics

The process of nuclear forensics includes the analysis and interpretation of nuclear material to determine its history, whether that be intercepted spent nuclear fuel, uranium ore concentrate, 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.

This study focuses on non-detonated materials, specifically, spent nuclear fuel. It is important to determine if some intercepted material is from a commercial fuel cycle or if it is meant for weapons production (and where the material was obtained from).

measure material, use isotope content and/or isotope ratios to determine things like reactor type, fuel type and enrichment at beginning of irradiation, cooling time, burnup. (Classification, Characterization, Interpretation (Analysis), Reconstruction (Attribution) - from the New Nuclear Forensics book) (Char methods to get isotopic ratios or use S/ML, Interp examples) After the material characteristics are measured, they are matched in a forensics database that includes some or all this information for pre-existing/pre-measured SNF. These databases are kept by individual countries, and a given database will have widely varying uncertainty depending where the material was measured as well as missing data in some fields. Therefore, matching can be difficult.

A lofty goal for the forensics community would be to develop methods that provide instantaneous information that is reliable enough to guide an investigation (e.g., within 24 hours). 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 spectroscopy 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. However, using a well-trained machine-learned model may be able to overcome these inherent issues with gamma spectra. The current and future work of this study is designed with this in mind.

Machine Learning

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 [2, 3, 4].

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 [1, 5, 6, 7, 8]. 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.

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.

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 work [1] while 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.

METHODOLOGY

herro

Training Set

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 [9]. Specifically, the ARP module of the activation and depletion code ORIGEN was used. [10]

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 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.

Model Comparison

RESULTS AND DISCUSSION

hello

CONCLUSION

oh hi

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship and the U.S. Department of Homeland Security's Nuclear Forensics Graduate Research Fellowship under Grant Award Number, 2012- DN-130- NF0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation or the U.S. Department of Homeland Security.

REFERENCES

1. K. DAYMAN and S. BIEGALSKI, "Feasibility of fuel cycle characterization using multiple nuclide signatures," *Journal of Radioanalytical and Nuclear Chemistry*, **296**, 195–201 (2013).
2. C. F. WEBER and B. L. BROADHEAD, "Inverse Depletion/Decay Analysis Using the SCALE Code System," in "Transactions of the American Nuclear Society Winter Meeting," Albuquerque, NM, USA (2006), vol. 95, pp. 248–249, track 4: Nuclear and Criticality Safety Technologies.
3. B. L. BROADHEAD and C. F. WEBER, "Validation of Inverse Methods Applied to Forensic Analysis of Spent Fuel," in "Proceedings of the Institute of Nuclear Materials Management 51st Annual Meeting," Baltimore, MD, USA (2010).
4. C. F. WEBER, V. A. PROTOPOPESCU, M. H. EHINGER, A. A. SOLODOV, and C. E. ROMANO, "Inverse Solutions in Spectroscopic Analysis with Applications to Problems in Global Safeguards," in "Proceedings of the Institute of Nuclear Materials Management 52nd Annual Meeting," Palm Desert, CA, USA (2011).
5. M. ROBEL, M. J. KRISTO, and M. A. HELLER, "Nuclear Forensic Inferences Using Iterative Multidimensional Statistics," in "Proceedings of the Institute of Nuclear Materials Management 50th Annual Meeting," Institute of Nuclear Materials Management, Tuscon, AZ, USA (Jul 2009), lLNL-CONF-414001.
6. M. ROBEL and M. J. KRISTO, "Discrimination of source reactor type by multivariate statistical analysis of uranium and plutonium isotopic concentrations in unknown irradiated nuclear fuel material," *Journal of Environmental Radioactivity*, **99**, *11*, 1789–1797 (Nov. 2008).
7. A. JONES, P. TURNER, C. ZIMMERMAN, and J. GOULERMAS, "Machine Learning for Classification and Visualisation of Radioactive Substances for Nuclear Forensics," in "Techniques and Methods for Safeguards, Nonproliferation and Arms Control Verification Workshop," Portland, Oregon (May 2014).
8. A. E. JONES, P. TURNER, C. ZIMMERMAN, and J. Y. GOULERMAS, "Classification of Spent Reactor Fuel for Nuclear Forensics," *Analytical Chemistry*, **86**, 5399–5405 (2014).
9. OAK RIDGE NATIONAL LABORATORY, "SCALE: A Comprehensive Modeling and Simulation Suite for Nuclear Safety Analysis and Design," Code suite, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA (Aug 2016), version 6.2.1, ORNL/TM-2005/39, Available from Radiation Safety Information Computational Center as CCC-834.
10. B. REARDEN and M. JESSEE, "Ch. 5 Depletion, Activation, and Spent Fuel Source Terms," in "SCALE Code System: User Documentation," Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA (Apr 2016), pp. 5–1–5–263, version 6.2.1; ORNL/TM-2005/39.