The Nuclear Forensics Problem and Statistical Methods: Evaluating Machine Learning for Spent Fuel Prediction

Arrielle C. Opotowsky, Paul P.H. Wilson

Computational Nuclear Engineering Research Group
University of Wisconsin at Madison
opotowsky@wisc.edu, paulwilson@wisc.edu

INTRODUCTION

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. To characterize the materials, both radiological methods (e.g., gamma spectroscopy) and ionization methods (e.g., mass spectroscopy) are used to determine isotopic ratios, chemical compounds, or trace elements. Although each category has a multitude of techniques within it, 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 determine the origin of the nuclear material(s). 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. Fortunately, machine learning algorithms can be explored to create a model from a database to "fill between the lines". Additionally, having a machine-learned model may overcome the above challenges of multidimensionality, missing data, and irregular uncertainty.

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 expensice. And although gamma spectroscopy can give extremely fast results cheaply, it only measures certain radiological signals and is influenced by many environmental (or storage) factors. In the simulation and machine learning paradigm, we need to determine what exactly is needed to train a machine-learned model. 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 and algorithms to use. This work is based off previous work on the subject (cite Dayman), and expands upon it by also evaluating a more advanced machine learning algorithm: neural nets. 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.

BACKGROUND AND THEORY

Nuclear Forensics

Nuclear forensics is the analysis and interpretation to determine the history of nuclear material, whether that be intercepted spent nuclear fuel, uranium ore concentrate, or the debris from an exploded nuclear device. Technical nuclear forensics focuses on the characterization and interpretation of those results. This, in combination with intelligence data, aids in material attribution, which is the overall goal of nuclear forensics.

Workflow for determining SNF quantities of interest: 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.

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.

(Classification, Characterization, Interpretation (Analysis), Reconstruction (Attribution) - from the New Nuclear Forensics book) (Char methods to get isotopic ratios or use S/ML, Interp examples)

To accomplish this, (we want to see if) it is possible to limit material measurements to rapid-result nondestructive analysis, such as gamma ray spectroscopy. Gamma spec gives less information at a higher uncertainty than the near perfect results of some destructive mass spec techniques, like TIMS. Additionally, within gamma spec techniques (field vs lab), uncertainties can also vary.

(Something on goal to get as much info as possible within 24 h of interception or device detonation).

Machine Learning

Given imperfect data with varying amounts of uncertainty as well as the required comparison to imperfect databases, many have begun considering artificial intelligence approaches to nuclear forensics problems, such as implementing searching algorithms for database comparison and machine learning for determining spent fuel characteristics (cite all). A variety of statistical and machine learning tools have been used to both classify spent fuel (reactor type, fuel type) and predict values such as burnup, initial enrichment, or cooling time (regression).

Add some real ML background here

EXPERIMENTAL DESIGN

This work begins by replicating the training and test sets used in ref (cite Dayman). The study does evaluate the imapact of randomly introduced error of varying amounts on the ability of the algorithms to correctly predict the burnup and reactor type. It also varies the amount of nuclides given as training to the algorithms (a full list, a short list determined by PCA, and fission products only). However, machine learning algorithms are heavily dependent on the inputs and parameters given to them, such as training set sizes, learning rates, etc. Additionally, it does not consider more advanced aglorithms. This paper adds in the comparison of the simpler algorithms to neural nets. It also presents learning curves for each algorithm to investigate the effect of the training set size on prediction error. For the classification of reactor type, ROC curves are also presented to evaluate prediction and generalization strength. For the regression of burnup, ??? pairwise t test?. Thus, I will first investigate these algorithms with no error introduced, but will extend the validation from a predetermined test set to be randomly partitioned cross validation. And the next study will introduce error by limiting the nuclides to only those that can be measured with a gamma spectrometer (future work).

Algorithms Used

NN classifier+regressor, ridge classification + regression, neural net classification + regression- these are all discussed above, but can talk deets + parameters here

Validation of Each Algorithm

Using predetermined test set:

1. Learning curves: For a given (randomly chosen) training set size for each algorithm, run several trainings and predictions. This helps determine if we are over or under training and each algorithm's robustness to over/under fitting.

2. Classification Training Error: Precision/recall values / ROC curves for the 3 classes or maybe comparing algorithms to each other for each class

3. Regression Training Error: Confidence intervals on predictions to understand true error versus sample error Test set must be > 30 instances, Can easily calculate N% confidence interval.

Using cross-validation instead:

1. redo learning curve study without having to run several trainings and predictions for each set (bc that's already done!)

2. redo precision/recall study

3. redo confidence interval study

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.

RESULTS AND ANALYSIS

So far, should include, with all 4 algorithms shown in each plot: 1. Learning curves + CV curve 2. ROC plots + CV plot 3. Confidence interval tables + CV table

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. Future work would be to investigate the number of cross validation folds.

ACKNOWLEDGMENTS

This material is based upon work supported a Department of Energy Nuclear Energy University Programs Graduate Fellowship.

REFERENCES