© 2018 Mark Kamuda

It's possible this word doesn't add any information to the title

AUTOMATED ISOTOPE IDENTIFICATION ~~ALGORITHM~~ USING ARTIFICIAL NEURAL NETWORKS

BY

MARK KAMUDA

THESIS preliminary examination

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Nuclear Plasma and Radiological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

~~Master's~~ Committee:

Professor Kathryn Huff, Adviser
Professor Rizwan Uddin
Professor Tomasz Kozlowski
Professor Mark Hasagawa-Johnson
Professor Clair Sullivan

ABSTRACT

Abstract goes here.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

question mark?

The main question addressed in this work is: Can artificial neural networks (ANNs) using simulated spectra as training data automate isotope identification and quantification. ANNs can incorporate abstract features of a gamma-ray spectrum in non-intuitive ways. This may give ANNs the ability to identify and quantify isotopes using large isotope libraries practical for domestic nuclear security, operate using low-resolution NaI radiation detectors without knowing the detector calibration or background spectrum.

comma splice

really, this doc is proposing dissertation work.

This dissertation will demonstrate the performance of ANNs for tasks related to identifying and quantifying isotopes in low-resolution gamma-ray spectra. The low-resolution detector of interest in this work is a 2-inch by 2-inch NaI(Tl) cylindrical scintillation detector. This detector is industry standard due to its ease of use, low cost, and acceptable resolution for gamma-ray spectroscopy. Tasks ~~that will be~~ investigated where will focus on identifying isotopes in the ANSI N42-34-2006 required list [1]. This ANSI standard also requires isotope identification algorithms to operate when the radioactive material is behind shielding. Accordingly, the impact of shielding will also be incorporated into this work.

In addition to isotope identification, this work will explore the ability of ANNs to quantify the count contribution from each isotope. The ANN's ability to extract count contribution information from gamma-ray spectra is important for nondestructive analysis (NDA). While NDA is useful for. The second purpose is a possible use of neural networks in post-detonation nuclear forensics. In post-detonation nuclear forensics, a large number of radioactive fission products are created. Quantifying the amount of isotopes in post-detonation debris can yield useful information about the device's properties. Mixtures of laboratory sources will be used as a surrogate for post-detonation debris and the ability for the ANN to accurately calculate the mixture components will be established.

This sentence is missing something

Another case where knowing the isotope quantities in a mixture of isotopes is in uranium enrichment calculations. Knowing the enrichment of uranium is particularly important in two areas. The first are is when uranium is identified at a border crossing. Typically the spectrum would need to be given to a trained spectroscopist to quantify the enrichment, but this process could be automated on the device first used to identify it. The second case is ~~would~~ be in treaty verification technologies. Low-resolution NaI gamma-ray detectors decrease the amount of possibly sensitive information collected while giving enough information to produce accurate enrichment quantification. The fact that ANNs can be taught to ignore certain patterns and only give information agreed upon by treaty signatories also makes them a good tool for treaty verification. The ability of ANNs to operate without knowing the shielding and background spectrum makes them better for zero knowledge scenarios. NaI detectors also have a higher efficiency than the higher resolution HPGe. This means that the counting times for NaI are smaller than would be required to get the same number of counts using an HPGe.

scenarios

with automated isotope identification

This is a tautology

CHAPTER 2

LITERATURE REVIEW

2.1 Gamma-Ray Spectroscopy for Isotope Identification

"challenges" perhaps?

There are a number of difficult things associated with performing isotope identification. Traditionally, isotope identification is conducted by a trained spectroscopoist. Once a spectrum of interest is identified, ~~a~~ the spectroscopist will identify photopeaks [2]. and thereby identify the isotope(s) present

There are many automated radioisotope identification methods available, but few perform well given a low-resolution gamma-ray spectrum of a mixture of radioisotopes. Common methods include library comparison algorithms, region of interest (ROI) algorithms, principle component analysis (PCA), and template matching.

Library comparison algorithms attempt to match photopeak energies found in a gamma-ray spectrum with those found in a library of known isotope decay energies. Drifts and uncertainties in detector calibration can lead to misidentifying photopeaks, leading to incorrect isotope identifications [3]. To be automated, this method needs an algorithm to extract photopeak centroids from a spectrum. Photopeak extraction algorithms face difficulties when a large number of photopeaks overlap in a spectrum, such as when a mixture of radio-isotopes are measured with a low-resolution detector [4].

ROI algorithms search for elevated counts compared to background in a region where photopeaks are expected to be for different radioisotopes. ROI algorithms may also operate poorly when photopeaks of different radioisotopes overlap [3]. For this reason, large isotope libraries will preform poorly using this method. Similarly to the library comparison algorithm, calibration drift may shift photopeaks into different neighboring ROIs, leading to incorrect identification. The ROI method has been used to differentiate normally occuring radioactive material (NORM) from special nuclear material (SNM) using plastic scintillators

[5].

PCA can also be applied to radioisotope identification. The goal of PCA is to reduce the dimensionality of a dataset into uncorrelated variables [6]. Using a few of these principle components, the data may be represented in a reduced space that contains most of the information present in the original data. The transformed data can then be clustered based on isotope identity. Clustering algorithms may include K-means or Mahalanobis distance [7, 8]. PCA has been applied to isotope identification using plastic scintillators [9] and anomaly detection using both plastic scintillators and NaI detectors [10]. Despite the progress of PCA in some isotope identification problems, there has not been significant progress in applying PCA to separating mixtures of isotopes in gamma-ray spectra.

Template matching algorithms find an example in a database of gamma-ray spectra that most closely matches a measured spectrum [3]. The database of spectra can contain multiple detector calibration settings, shielding materials, and source-to-detector distances. Goodness of fit can be measured using a hypothesis test such as chi-squared test, euclidean distance, or Mahalanobis distance. While a sufficient amount of example spectra can be used to identify almost any measured spectrum, the drawback of this method is the time necessary to compare a measured spectrum to the library and the computer memory necessary to store said library. This method also may have difficulty when mixtures of isotopes are considered, although work is being done to correct this [11].

2.2 Artificial Neural Networks

Artificial neural networks were first created to mimic biologic neurons. Since their creation, they have demonstrated promising results on a variety of different classification and regression tasks [12, 13, 14]. The following sections will give an overview of how ANNs learn and operate.

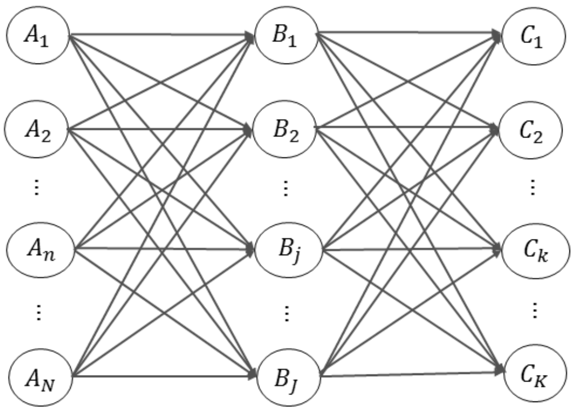


Figure 2.1: Example ANN with input layer A, hidden layer B, and output layer C.

2.2.1 Architecture and Training

ANNs work by mapping arbitrary input spaces, $\mathbb{R}^N$, to arbitrary output space, $\mathbb{R}^K$. An example one hidden layer ANN mapping $\mathbb{R}^N \mapsto \mathbb{R}^K$ is shown in Figure 2.1. Each circle represents a neuron, or node. The mathematical process governing each neuron in the ANN is shown in Figure 2.2. In Figure 2.2, the signal from the previous layer is propagated to the next by applying some function, typically sigmoidal, to the dot product of the signal from the previous layer and the weight vector going into a given node, B_j in Figure 2.2. Given a one-layer ANN with a finite number of hidden nodes, any function $\mathbb{R}^N \mapsto \mathbb{R}^K$ can be described to arbitrary precision [15]. Additional hidden layers increase the representational power of an ANN, reducing the number of nodes and computational power required to represent a function. There is no direct method to compute the number of hidden layers or nodes for a given problem. These, along with other hyperparameters, need to be optimized for a given dataset.

perhaps this deserves its own sentence

appropriate/ ideal/ optimal

Artificial ANNs learn a function by changing the weights connecting the layers so that some error function is minimized for a given dataset. One popular method to update the weights is through the process of gradient descent through the backpropogation of errors [16]. The update equation for a single weight, w_j, is shown in Equations 2.1 and 2.2. In these equations, *Error* is the given error function to be minimized (commonly mean squared error or cross entropy) and η is the learning rate of the ANN.

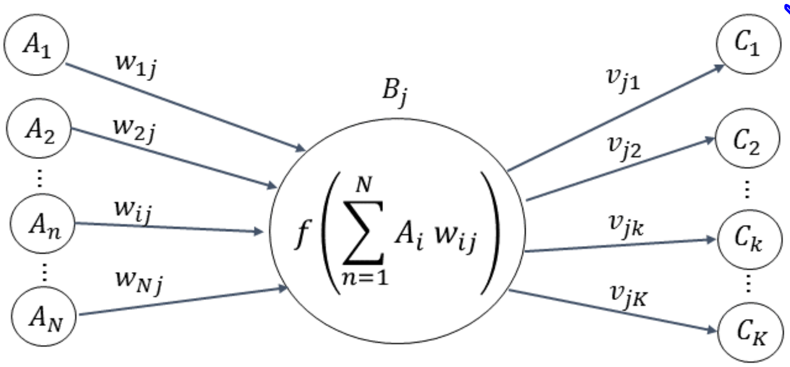


Figure 2.2: Summary of the operation of a single neuron.

did you create this diagram? If yes, great. If not, cite the source

$$\Delta w_j = -\eta \frac{dError}{dw_j} \tag{2.1}$$

could we use a symbol (or just E) to represent error?

$$w_j^{new} = w_j^{old} + \Delta w_j \tag{2.2}$$

2.2.2 Hyperparameters

In general, ANNs have a tendency to memorize their training set in a process called overtraining. An overtrained ANN will tend to incorrectly identify novel data. To prevent this, a number of hyperparameters were used to prevent overfitting and optimize performance. Unfortunately, there is currently no known method to know which hyperparameters have an impact on model performance before training. Because of this, a number of popular hyperparameters are typically added to a model and a random hyperparameter search is used to identify those which are important. There is evidence that a random search in a given hyperparameter range finds better hyperparameters quicker than a grid search in the same range [17]. There is also a proof showing that given 60 points randomly sampled in some space with a finite minimum, the minimum of those 60 random samples is within 5% of the true minimum with 95% probability [18]. Training 60 ANNs is computationally feasible, making this a good method to find a close-to-optimal ANN for a given dataset.

2.3 Isotope Identification Using ANNs

There have been a number of published papers which apply ANNs to automated isotope identification. ANNs have been applied to peak fitting [19], isotope identification [20, 21], and activity estimation [20, 22]. Many of this work rely on ROI methods [23], feature extraction [24], high-resolution gamma-ray spectra as the input to the ANN [25], small libraries of isotopes, and assume perfectly calibrated detectors. ANN training methods created for high-resolution gamma-ray spectra may not perform well when trained using low-resolution spectra given the large discrepancy in resolution. ANN training that relies on ROI methods may not perform well when ROIs overlap significantly with large libraries of isotopes.

It has been shown that an ANN may be trained to perform isotope identification and quantification using low-resolution NaI gamma-ray spectrum using a library of five isotopes [26]. While promising, this study did not include complicated source mixture analysis. This study also used a library too small to be of practical use. The American National Standards Institutes (ANSI) has identified 31 gamma-ray emitting isotopes that automated isotope identification algorithms should be able to identify [1].

did they includ simple Source mixture analysis?

And how many were identified in the too small Study?

CHAPTER 3

ARTIFICIAL NEURAL NETWORK APPROACH TO IDENTIFYING AND QUANTIFYING ISOTOPES IN GAMMA-RAY SPECTRA

3.1 Introduction

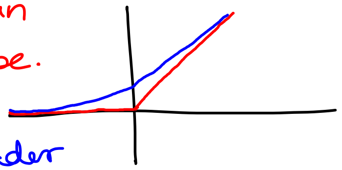
The ANN presented here is trained to quantify the count contribution of each isotope from a library. The ANN uses simulated gamma-ray spectra, this make simulating additional datasets simple and quick. The ANN structure, training details, and hyperparameter optimization are described.

3.2 Artificial Neural Network Structure

"uses a"
"uses the" ?
"used a"
"used the"

The fully connected ANN explored in this work use rectified linear (*relu*) activation function, seen in Equation 3.1, with a *softmax* output function, seen in Equation 3.2.

relu has an easy shape.



but the reader might be helped by a diagram showing softmax

$$relu(x) = max(0, x). \tag{3.1}$$

only one relu function...

$$softmax(z_j) = \frac{\exp(z_j)}{\sum_{k=1}^{K} \exp(z_k)}. \tag{3.2}$$

The *relu* function was chosen for the node activations because they are easy to optimize and generally perform better than other non-linear functions [27, pg. 189]. While the *softmax* function is traditionally used in classification ANNs, using it here ensures the output from the ANN is normalized to unity. This allows the ANN to output relative count contributions from each isotope. This method was shown to perform well for quantifying isotopes in real and simulated spectra [28, 29]. By setting a detection threshold based on the highest count contribution, this method can also be used as an identification algorithm. For

This sentence is missing something. Reads like a tautology. Related: probably worth explaining why 15% was chosen.

example, Table 3.1 shows the top five isotope outputs from an ANN that used the spectrum in Figure 3.1 as input. Using a threshold of 15%, any isotope with a contribution below 15% of 0.441 can be ignored. This leads to a correct identification of ^{60}Co, ^{137}Cs, and background.

Table 3.1: Top five isotopes found by an ANN trained to quantify isotopes.

Isotope	Count Contribution
^{60}Co	0.441
^{137}Cs	0.440
Background	0.068
^{99}Mo	0.019
^{235}U	0.017

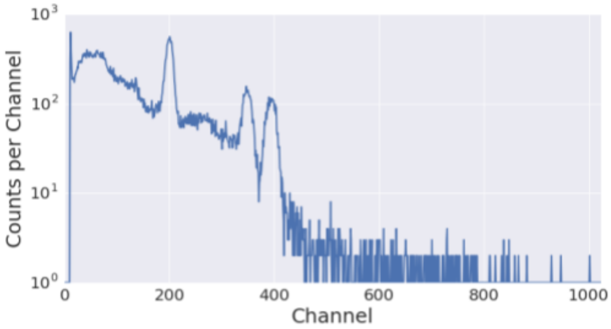


Figure 3.1: Gamma-ray spectrum of a 0.288 μCi ^{60}Co source measured from a distance of 7.5 cm from the detector face and a 0.890 μCi ^{137}Cs source. The count rate on the detector face is

3.3 Dataset Construction

In order to train an ANN, a training set and training key must be provided. The training set is a set of ANN input data and the training key is the correct ANN output for each input. Because creating a training set of real gamma-ray spectra is infeasible, the training set used in this work was simulated. The training set was created using a one-dimensional Monte Carlo radiation detector simulation program called GADRAS [30]. The simulation process began by simulating individual 1 mCi sources with no background. Each source was simulated at a distance of 30cm from a 2 inch by 2 inch Ortec 905-3 NaI spectrometer for 10 hours, ensuring each spectrum had low statistical noise. These single isotope sources were then sampled using the inverse transform sampling method [31]. This allows the creation of arbitrary source combinations. The background isotopes were modeled using built-in GADRAS background sources for ^{40}K, uranium and daughters in soil, and thorium and daughters in soil.

random?

Because the calibration on NaI detectors shifts over time (due to voltage drift in the electronics and changes in crystal temperature), a method to change training spectra calibration was included. To mimic gain shift, the channels in each spectrum were linearly rebinned by some percent. After rebinning, the resulting spectrum was reconstructed using third order spline interpolation with the new bin positions.

Using this method, many training sets can be created depending on different algorithm goals. The training set and isotope library can be modified for specific problems like unknown source interdiction, uranium enrichment calculations, and post-detonation nuclear forensic debris analysis. Methods of creating ANNs for these problems will be described in later sections.

expand acronyms on first use

3.4 Training Details and Hyperparameter Optimization

citation for superior performance

Once the ANN hyperparameters are decided on, an optimization algorithm is needed to train the model. The ADAM optimizer [32] was chosen as the training algorithm for this work due to its incorporation of parts of other successful optimization algorithms and its reported superior performance over these algorithms. Another benefit of the ADAM optimizer is introduction of only one additional hyperparameter, the learning rate. Other optimizers require tuning more than one additional hyperparameter. During training, The ~~cost function the~~ ANN minimized ~~during training was~~ the average cross entropy between the correct labels, y_n, and the network predictions, $\hat{y}_n$, seen in Equation 3.3. This cost function was chosen because it is traditionally used with ANNs whose output is the *softmax* function. Why is that?

$$E = -\frac{1}{N}\sum_{n=1}^{N} y_n \log(\hat{y}_n) + (1-y_n)\log(1-\hat{y}_n). \tag{3.3}$$

Ideally, the equation should be a part of a sentence

Raw data is typically preprocessed before being input to an ANN. For this work each spectrum was preprocessed by scaling the counts in each bin between [0,1]. Scaling the inputs improves numerical stability during training.

It is possible this should be explained before the ADAM optimizer.

As previously discussed, hyperparameters are often necessary to properly train an ANN. The following hyperparameters are considered in this study are: the number of neurons in each layer, the number of layers used, initial learning rate for the training algorithm, the L_2 weight regularization strength, and neuron dropout rate.

Adding L_2 weight regularization allows the magnitude of the weights to increase only when there is a comparable reduction in the unmodified error function.

$$\tilde{E} = E + \sum_i \lambda w_i^2. \tag{3.4}$$

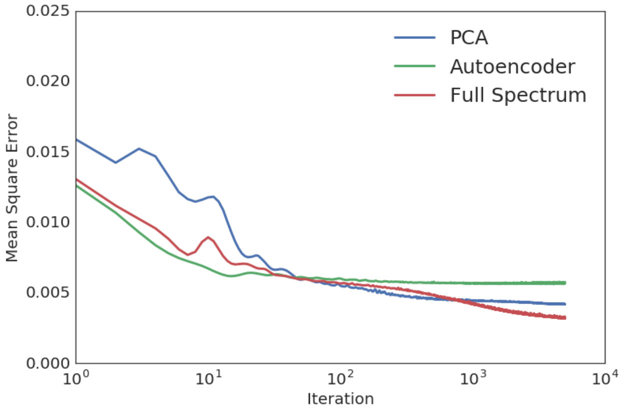
In Equation 3.4, w_i is the weight between each neuron in the ANN and λ is the regularization strength hyperparameter. A larger λ will force the ANN to prefer smaller weights connecting the neurons. If λ is too small, the model is more likely to overfit. If λ is too large, the ANN will preferentially minimize the L_n error, failing to learn the desired task.

have we defined this?

Another method to reduce model capacity is neuron dropout. Neuron dropout is the process of temporarily removing a neuron from the ANN architecture during training [33]. The probability that a neuron is removed is called the neuron dropout rate, which is a hyperparameter. By applying dropout throughout training, the ANN's architecture changes every iteration. The makes neuron dropout a cost efficient way to average many different ANN architectures, improving performance.

3.5 Key Results

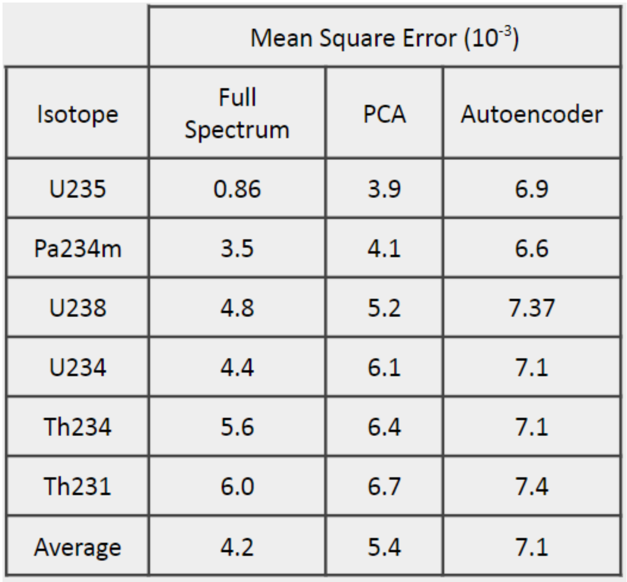
explanation in text.



comparing method for input model reduction.

d

Figure 3.2: Mean square error vs training iteration for three different ANN inputs.



	Mean Square Error (10^{-3})		
Isotope	Full Spectrum	PCA	Autoencoder
U235	0.86	3.9	6.9
Pa234m	3.5	4.1	6.6
U238	4.8	5.2	7.37
U234	4.4	6.1	7.1
Th234	5.6	6.4	7.1
Th231	6.0	6.7	7.4
Average	4.2	5.4	7.1

Figure 3.3: Table showing the average mean square error from 1000 simulated enriched uranium spectra.

CHAPTER 4

FUTURE WORK AND PROPOSED EXPERIMENTS

For the proposed work

will

There are a number of experiments that would build on the work that has been done. These experiments are described in this section.

4.1 Proposed Datasets

4.1.1 Unknown Source Interdiction

Sentence missing a word?

One important problem in automated gamma-ray spectroscopy giving untrained operators the ability to identify unknown hidden sources using hand-held devices. ~~There are~~ M many issues with this problem~~, but they~~ can be addressed by properly constructing a training set for an ANN.

"nefarious"? "covert"? "hidden"?

Typically these sources produce weak signals due to being purposely shielded. Because lower-energy gamma-rays are preferentially attenuated over higher-energy gamma-rays, shielding also changes the shape of a gamma-ray spectrum.

flip the order of this sentence for clearer causality?

In addition to the previous hurdles, because untrained operators would be using these algorithms, the detector calibration cannot be completely trusted.

To address the problems described above, the training dataset will be constructed of simulated spectra of

While the possible threat source is unknown, the isotopes in the ANSI N42-34-2006 have been identified as Source strengths will range from μCi to a Ci and source-to-detector distance will range from 10cm to 1 meter. Sources below a μCi do not produce sufficient counts to be detected. Sources above a Ci are

The performance of the ANN on this dataset will be reported using a ROC curve for a

expand acronym on first use.

consider glossaries package

simulated spectra dataset as well as a measured spectra dataset. The simulated spectra dataset will be composed of lab isotopes behind various shielding materials.

4.1.2 Post-Detonation Nuclear Forensic Debris Analysis

Immediately after a nuclear detonation, first responders may ~~be~~ collect~~ing~~ a large number of gamma-ray spectra using handheld low-resolution detectors. These detectors may have an unknown or poor calibration and each spectrum may be measured for a short amount of time. Despite these drawbacks, the data is still valuable because it can be used to determine isotopics of the debris generated by the explosion [34]. Due to the complicated gamma-ray spectrum produced by a large number of radioactive fission products, many photopeaks and spectral features will overlap in a spectrum. This feature overlap increases the difficulty of and slows photopeak analysis, especially for low-resolution spectrometers.

→ So, you'll try the ANN on post-det spectra?

4.1.3 Uranium Enrichment Calculations

For treaty verification purposes, quickly measuring the enrichment of uranium is important. It can be argued that HPGe detectors are better suited for this task due to their higher-resolution over NaI.

To test the ability for ANNs to measure uranium enrichment,

4.2 Autoencoders

perhaps a definition which doesn't rely on the word "encoding" will be more informative.

An autoencoder is an unsupervised ANN whose goal is to learn an encoding of the input. This is accomplished by simultaneously training an encoding ANN and a decoding ANN. The encoding ANN reduces an $n-$dimension signal to a $m-$dimension signal, where $m < n$. The decoding ANN takes the $m-$dimension signal and attempts to reproduce the $n-$dimension input or one similar.

Figure would be helpful here

In previous work, a single ANN had to learn multiple tasks to identify isotopes. An ANN would have to simultaneously identify the detector calibration, background signal, and

→ Citation

possible source signal. By training an autoencoder to reconstruct a background subtracted and correctly calibrated spectrum, the task of isotope identification is simplified for the ANN. This may result in more accurate identifications. To test this, a single autoencoder and three ANNs will be trained. The first will be trained without an autoencoder. The second will be trained using the encoder as input. The third will be trained using the full autoencoder as input. A manual search or a random hyperparameter search will be used to find an appropriately structured autoencoder. The testing and validation error for these ANNs will be compared for each dataset described previously.

*

distinction between these is unclear.

In addition to using fully connected autoencoders, a 1-D convolutional autoencoder will also be explored. Fully connected ANNs do not assume the input has local spatial structure, while convolution ANNs do. Because gamma-ray spectra have local spatial structure in the form of photopeaks and Compton continua, it may be better to use a convolutional ANN over a fully connected ANN.

Good! Maybe explain how?

4.3 Additional Simulated Detector Models

Simulating addition NaI detectors using GADRAS may help the ANN generalize to different detectors. To begin this process, two 2-inch by 2-inch NaI detectors will be modeled using GADRAS and their properties will be changed (for example: crystal dimension, calibration, scattering environment). The number of different detectors modeled will be increased and the generalization performance of the ANN will be evaluated. The generalization performance can be calculated by checking the ANNs performance on spectra produced by a detector whose properties the ANN has not seen during training.

This is good. Perhaps add a table of the parameters & the ranges over which you'll vary them.

4.4 Mixture Validation Dataset

A real dataset of real spectra are needed to confirm the models performance. Mixtures of laboratory isotopes can be made. Mixtures will vary by signal-to-background ratio by varying source-to-detector distance and integration time. Isotopes available are: ^{60}Co, ^{137}Cs, ^{133}Ba, ^{152}Eu, ^{22}Na and others.

Will these be appropriate for validation? Do any have overlapping peaks? (elaborate)

4.5 Bagging

Each time the ANN trains, it produces slightly different identification results and performance. Bagging (bootstrap aggregating), or the process of averaging the outputs of many ANNs, can reduce the variance in the output [35]. In addition to this, bagging more accurately displays the true performance of a given ANN structure. The number of ANNs included in the bagging process will be explored.

demonstrates

To find the optimal number? How will you know the optimum has been reached? What metric will you use?

4.6 K-folds cross validation

Perhaps a diagram of this whole data workflow should be used throughout this document. (i.e. Is this before/after autoencoding? Which data will train it? ...)

K-folds cross validation will be incorporated into the model. Cross validation is the general method of determining how well a model will generalize to novel data. K-folds cross validation does this by splitting the available data into k subsets. From these, k-1 subsets are used to train the ANN and the remaining subset is used as the validation dataset. This process is repeated, using each subset once as the validation dataset. Typical values for k are either 5 or 10. Cross validation more accurately analyze how well a give ANN structure and set of hyperparameters will perform. K-folds cross validation will be added to hyperparameter optimization. The error and hyperparameter structure will be compared for ANNs trained with no cross validation, 5-folds cross validation, and 10-folds cross validation.

4.7 Latin Hypercube Sampling for Training Set Construction

Currently, the training set is constructed using random combinations of isotopes with random count rates. Because random sampling does not guarantee the training space is well sampled and often produces clusters, this may not be a good method to construct the training dataset. This may lead the ANN to learn a bias for random clusters in the training set. A way to reduce the chances of clusters and more uniformly sample the input space is using Latin hypercube sampling (LHS). LHS partitions each dimension of a space into N equal parts. The space is then sampled using N points, ensuring that there is only one point along each partition in each dimension.

LHS can be vizualized nicely.

To quantify how much LHS helps identification for a given data space, ANNs will be trained with an increasing number of samples. The ANNs performance on a real dataset will then be measured and compared. It is expected that at a certain number of samples the ANN performance will reach an asymptote. The number of samples may change for each data space (no shielding, shielding, uranium enrichment) depending on the complexity of the problem.

a table of these parameters and estimates of the ranges over which you will vary them is appropriate.

4.8 Model Confidence Using the Dropout Uncertainty Method

We can exploit the dropout method used to regularized the ANN to get a confidence measure given a spectrum [36]. Once the ANN is trained, an unknown sample is presented to the model and the solution is recorded. Dropout is then used on the model using the same dropout probability that was used to train the model. The same unknown sample is then passed through the ANN and the answer recorded. This process of dropout and recording the resulting output is repeated. The variance of the outputs from this process can be used to determine the ANNs confidence in a given pattern.

this text description is tricky to follow. Perhaps a corresponding diagram — would clarify

Vague.

A few results are predicted using the ANN confidence. It is also expected that high signal-to-background measurements and isotope mixtures of a few components will have a high model confidence. It is expected that low signal-to-background measurements and mixtures of many components will have a low model confidence. Additionally, the model confidence with an isotope not included in the training set should be very low. This confidence measure may be used to reduce the false alarm rate when performing isotope identification.

4.8.1 New Training Stopping Condition

ANNs need a condition to end training. This condition is typically when a certain allowable error in a validation dataset is reached or when the training error does not appreciably decrease over time. Previously in this work, the stopping condition has been based on the cost function the ANN is trying to decrease, the cross entropy. It may be better to stop training when a more useful metric. ~~An example of this metric is~~ when the maximum error

With?

For example, training could complete

in the training set reaches a certain threshold. Another metric is to stop training when the ANNs mean squared error for a validation set stops decreasing.

Such a threshold might be problem specific. How might you select this threshold?

CHAPTER 5

CONCLUSION

Previous work has shown that ANNs are capable of solving problems in gamma-ray spectroscopy. Investigating more advanced ANN methods and proposed datasets may improve on previously reported performance.

CHAPTER 6

REFERENCES

[1] *American National Standard Performance Criteria for Hand-Held Instruments for the Detection and Identification of Radionuclides, ANSI N42.34-2006*, IEEE Std., Rev. ANSI N42.34-2006, 2007.

[2] M. Rawool-Sullivan, J. Bounds, S. Brumby, L. Prasad, and J. Sullivan, "Steps toward automated gamma ray spectroscopy steps toward automated gamma ray spectroscopy: How a spectroscopist deciphers an unknown spectrum to reveal the radioactive source," 02 2010.

[3] T. Burr and M. Hamada, "Radio-isotope algorithms for NaI gamma spectra," *Algorithms*, vol. 2, pp. 339–360, March 2009.

[4] H. Xiong, "Automated wavelet analysis of low resolution gamma-ray spectra and peak area uncertainty," Master's thesis, University of Illinois at Urbana-Champaign, 2015.

[5] J. Ely, R. Kouzes, J. Schweppe, E. Siciliano, D. Strachan, and D. Weier, "The use of energy windowing to discriminate SNM from NORM in radiation portal monitors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 560, no. 2, pp. 373 – 387, 2006.

[6] J. I. T., *Principal Component Analysis.* Springer New York, 2002.

[7] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, Jul 2002.

[8] V. V. Kumari, B. R. Raju, and A. Naik, "Hybrid clustering algorithm based on mahalanobis distance and mst," *International Journal of Applied Information Systems*, vol. 3, no. 5, pp. 60–63, Jul 2012.

[9] D. Boardman, M. Reinhard, and A. Flynn, "Principal component analysis of gamma-ray spectra for radiation portal monitors," *IEEE Transactions on Nuclear Science*, vol. 59, no. 1, pp. 154–160, Feb 2012.

[10] R. Runkle, "Analysis of spectroscopic radiation portal monitor data using principal components analysis," *IEEE Transactions on Nuclear Science*, vol. 53, no. 3, pp. 1418–1423, 2006.

[11] J. Mattingly and D. Mitchell, "A framework for the solution of inverse radiation transport problems," *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, pp. 3734–3743, 2010.

[12] S. Jeyanthia, N. Maheswaria, and R. Venkatesh, "Neural network based automatic fingerprint recognition system for overlapped latent images," *Journal of Intelligent Fuzzy Systems*, vol. 28, 2015.

[13] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Annual Conference on Neural Information Processing Systems*, pp. 1097–1105, 2012.

[14] A. Rababaah and S. D., "Integration of two different signal processing techniques with artificial neural network for stock market forecasting," *Academy of Information Management Sciences Journal*, vol. 18, pp. 63–80, 2015.

[15] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. 4, pp. 251–257, 1991.

[16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by backpropagating errors," *Nature*, vol. 323, pp. 533–536, 1986.

[17] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.

[18] A. Zheng, *Evaluating Machine Learning Models.* O'Reilly Media, Inc., 2015.

[19] R. Abdel-Aal, "Comparison of algorithmic and machine learning approaches for the automatic fitting of gaussian peaks," *Neural Computing and Applications*, vol. 11, no. 1, pp. 17–29, 2002.

[20] R. Abdel-Aal and M. Al-Haddad, "Determination of radioisotopes in gamma-ray spectroscopy using abductive machine learning," *Nuclear Instruments and Methods in Physics Research A*, vol. 391, pp. 275–288, 1996.

[21] M. Medhat, "Artificial intelligence methods applied for quantitative analysis of natural radioactive sources," *Annals of Nuclear Energy*, vol. 45, pp. 73 – 79, 2012.

[22] V. Vigneron, J. Morel, M. Lepy, and J. Martinez, "Statistical modelling of neural networks in y-spectrometry," *Nuclear Instruments and Methods in Physics Research A*, vol. 396, p. 642647, 1996.

[23] V. Pilato, F. Tola, J. Martinex, and M. Huver, "Application of neural networks to quantitative spectrometry analysis," *Nuclear Instruments and Methods in Physics Research A*, vol. 422, pp. 423–427, 1999.

[24] L. Chen and Y.-X. Wei, "Nuclide identification algorithm based on KL transform and neural networks," *Nuclear Instruments and Methods in Physics Research A*, vol. 598, pp. 450–453, 2009.

[25] E. Yoshida, K. Shizuma, S. Endo, and T. Oka, "Application of neural networks for the analysis of gamma-ray spectra measured with a Ge spectrometer," *Nuclear Instruments and Methods in Physics Research A*, vol. 484, pp. 557–563, 2002.

[26] P. Olmos, J. C. Diaz, J. M. Perez, P. Gomez, V. Rodellar, P. Aguayo, A. Bru, G. Garcia-Belmonte, and J. L. Pablos, "A new approach to automatic radiation spectrum analysis," *IEEE Transactions on Nuclear Science*, vol. 38, pp. 971–975, August 1991.

[27] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* MIT Press, 2016, http://www.deeplearningbook.org.

[28] M. Kamuda, "Automated isotope identification algorithm using artificial neural networks," Master's thesis, University of Illinois at Urbana-Champaign, 2017.

[29] M. Kamuda, J. Stinnett, and C. J. Sullivan, "Automated isotope identification algorithm using artificial neural networks," *IEEE Transactions on Nuclear Science*, vol. 64, pp. 1858 – 1864, 2017.

[30] D. J. Mitchell and L. T. Harding, "GADRAS isotope ID user's manual for analysis of gamma-ray measurements and api for linux and android," *SAND2014-3933*, May 2014.

[31] L. Devroye, *Non-Uniform Random Variate Generation.* Springer-Verlag New York Inc., 1986.

[32] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference for Learning Representations*, 2015.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[34] K. J. Moody, P. M. Grant, and I. D. Hutcheon, *Nuclear Forensic Analysis.* CRC Press, 2014.

[35] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug 1996. [Online]. Available: https://doi.org/10.1007/BF00058655

[36] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," *International Conference on Machine Learning*, 10 2016.