Nuclear Detection Using Higher-Order Topic Modeling

Christie Nelson
RUTCOR
Rutgers University
Piscataway, New Jersey
cgrewe@eden.rutgers.edu

William M. Pottenger
Computer Science, DIMACS, RUTCOR
Rutgers University
Piscataway, New Jersey
drwmp@rci.rutgers.edu

Hannah Keiler
Statistics
Columbia University
New York, New York
hpk2108@columbia.edu

Nir Grinberg
Computer Science, DIMACS
Rutgers University
Piscataway, New Jersey
nirg@cs.rutgers.com

Abstract **In this paper, a novel approach to topic modeling based on the Higher Order Learning framework, Higher-Order Latent Dirichlet Allocation (HO-LDA), is applied to a critical issue in homeland security, nuclear detection. In addition, this research strives to improve topic models in the 'real time' environment of online learning. In total, seventeen different nuclear radioisotopes are classified, and performance of Higher-Order versus traditional techniques is evaluated.**

This project employs LDA and HO-LDA on a nuclear detection numeric dataset to gain a topic decomposition of instances. These learned topics are then used as features in a traditional supervised classification algorithm. In essence, the LDA or HO-LDA topic assignments are used as features in supervised learning algorithms that predict the class (isotope), treating LDA or HO-LDA as a feature space transform. Using Topic Modeling on numeric nuclear detection data is cutting edge, as to our knowledge this has never been done before on a nuclear detection dataset. Two methods of feature transformation are evaluated, including Multinomial Feature Creation and Maximum Channel Value Feature Creation. Results demonstrate further evidence that Higher Order Learning techniques can be usefully applied in topic modeling applied to nuclear detection.

Keywords: nuclear detection; higher-order learning; latent Dirichlet allocation; LDA; higher-order latent Dirichlet allocation; HO-LDA; topic modeling; higher-order topic modeling; machine learning;

I. INTRODUCTION

One of the challenges facing our world today is the proliferation of weapons of mass destruction, including the 'dirty bomb' (a conventional bomb that contains explosives as well as radioactive material). One of the technologies employed to detect nuclear materials usable in such devices are handheld radiation detectors. Such detectors employ various technologies for radio-nuclear classification based on the statistical properties of spectral emissions, which are high-dimensional in nature. When classifying such data, the most important task is to distinguish known or unknown threat isotopes from harmless radioisotopes. These isotopes must also be distinguished from the naturally occurring radioactive background. The ability to classify isotopes when the signal is not strong can be especially useful. In previous research, signal data from a handheld radiation detector named the Interceptor™ was studied in order to improve the detection and identification of nuclear isotopes at seaports, in particular with the Higher-Order Naïve Bayes (HONB) classifier [8]. The prior research focused on individual isotope detection to determine if an individual radioisotope was present or absent. The next step of this research effort is to explore the use of a new class of learning algorithms known as Topic Models in the domain of nuclear detection. Given the growing importance of modeling data in real-time, leveraging topic models using online learning is also an important focus of the research. In this context, we present a novel approach to topic modeling based on the Higher Order Learning framework, Higher-order Latent Dirichlet Allocation (HO-LDA), and its application to a critical issue in homeland security, nuclear detection.

This paper is organized as follows. In section II, background and related work is presented. Following this, the Approach and Results are presented in sections III and IV, followed by Conclusions in section V.

II. BACKGROUND AND PRIOR WORK

A. Nuclear Detection

The first atomic bomb was detonated in New Mexico in July 1945. Since then, nuclear warfare has been a top issue of national security. Currently, a major concern is nuclear terrorism. The fear is that a terrorist group will make, steal, or obtain a nuclear weapon or nuclear materials to produce a 'dirty bomb'. Despite the fact that this is an extremely unlikely scenario, the extraordinarily high consequences make it an important topic in national security. The detonation of a 'dirty bomb' is considered much more feasible (versus a terrorist group stealing or making a nuclear weapon), and this is a scenario that also could have a horrific outcome. While a dirty bomb wouldn't contain as high of an amount of radiation as a nuclear weapon, the bomb could still cause many deaths, and the residue from the blast could cause contamination of the area, though not nearly as much as from a nuclear weapon. Much of the damage from a dirty bomb would be from the initial blast; however the economic disruption caused by such a terrorist act is also a significant concern.

The use of scanners at seaports for the detection of nuclear isotopes is thus important for national security. Improvement of the performance of such scanners is an important technical goal, particularly handheld scanners such as the Thermo Scientific Interceptor™. The goal of this research is the improvement of the detection and identification of nuclear radioisotopes.

B. Topic Modeling

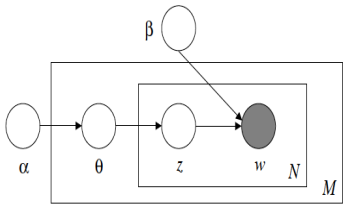
Topic Models are probabilistic models that model common patterns across a corpus of related text documents. Topic

models treat the documents as a set of observations that are generated by a probabilistic process using latent information, a set of topics. This latent information is then learned using posterior inference and can be used to perform inference on new data. There are several variations of topics models, each depicted by a directed graphical model. An important common underlying assumption of these models is that the documents are a bag-of-words, in other words, the order or words is not important. Latent Dirichlet Allocation (LDA) is seminal work which was extended in several follow-on research efforts.

LDA was first proposed by Blei in [9]. In LDA, the observed data are the words of each document and the hidden variables represent the latent topical structure, in other words, the topics themselves and how each document exhibits them. Given a collection, the posterior distribution of the hidden variables given the observed documents determines a hidden topical decomposition of the collection. Applications of topic modeling use posterior estimates of these hidden variables to perform tasks such as information retrieval and document browsing.

The following is the generative process to generate a corpus of documents using the Dirichlet Model per Figure 1:

Choose $\theta \sim$ Dirichlet(α)
For each of the N words w_n:
Choose a topic $z_n \sim$ Multinomial(θ)
Choose a word w_n from $p(w_n \mid z_n; \beta)$, a



multinomial probability conditioned on the topic z_n and parameterized by the topic distributions β

Figure 1. General Process to generate a corpus of documents using the Dirichlet Model

Inference in most topic models is performed using variational inference algorithms. These are heuristic algorithms that have the advantage of being very fast as compared to the traditional Gibbs Sampling algorithms. In order to perform inference using a Gibbs Sampling algorithm, the conditional probability of occurrence of a topic for a word in the corpus (given the topic labels of the words) is used. Since Dirichlet is the conjugate prior to the multinomial distribution, this probability turns out to have the following elegant closed form:

$$P(z_i = j|\mathbf{z}_{-i}, \mathbf{w}) \propto \frac{n_{-i,j}^{(w_i)} + \beta}{n_{-i,j}^{(\cdot)} + W\beta} \frac{n^{(d_i)} + \alpha}{n_{-i,\cdot}^{(d_i)} + T\alpha}. \quad (1)$$

This approach is used for sampling the topic for the term w at position i. The term $n_{-i,j}^{(w)}$ corresponds to the number of occurrences of the term w that are assigned to the topic j, not including the current (i^{th}) occurrence. Intuitively, the above formula can be interpreted as a word being assigned to a topic proportional to its frequency of occurrence in that topic.

One important aspect of LDA is that it does not directly model correlation between the occurrence of topics. In many real corpora, it is natural to expect that topics are correlated. For example, quantum mechanics and linear algebra are more likely to occur together in a document rather than pharmacy and linear algebra. LDA does not model this behavior mainly because of the use of the Dirichlet distribution. Topic modeling algorithms that model topic correlations directly lack the mathematical simplicity of LDA, and as a result require more complex iterative solvers.

C. *Higher Order Learning*

Higher Order Learning is an approach that utilizes relationships between attributes and features across instances. Traditional machine learning techniques assume that instances are IID (independent and identically distributed). Such traditional methods can be called "zero-order" because they do not leverage relationships between attributes. The traditional IID assumption does not permit traditional machine learning methods to leverage such "higher-order" relationships.

As shown in Figure 2 below, Higher Order Learning utilizes relationships between attribute values across instances. In Figure 2, there are three sample instances shown, instances V_1, V_2, and V_3. Instance V_1 has two attributes, attributes x_1, and x_2, instance V_2 has two attributes (x_2 and x_3), and instance V_3 has two attributes (x_3 and x_4). Traditional machine learning does not leverage the latent higher order paths. However, Higher Order techniques use these higher order paths to create a link between attributes. In this example, attributes x_1 and x_4 are linked by leveraging the higher-order paths between attributes V_1, V_2, and V_3.

One of the thrusts of this research is to evaluate if higher-order techniques can be successfully used in topic modeling algorithms, applied to the detection of nuclear radioisotopes. Higher Order Learning has already been shown to work well in several other applications in the statistical relational learning field.

D. *Problem Definition*

The focus of this research is the evaluation of the use of Higher Order Learning in topic modeling. The domain for evaluation is the detection and identification of nuclear materials. Given nuclear detection data, harmless radioisotopes must be distinguished from those that are potentially harmful. Thus, algorithms that can classify nuclear signals are desirable. The purpose of this research is thus to demonstrate the validity of leveraging statistical relational information in topic models

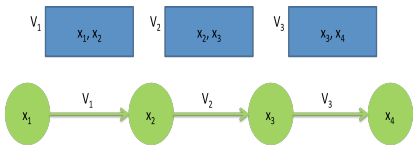


Figure 2. Illustration of a higher-order path (V_i are instances, x_i are attributes)

in the domain of nuclear detection.

E. *Prior Work*

Prior work was conducted by the second author of this paper and Jason Perry on the investigation of the nuclear detection dataset [3]. The original data was studied in detail in order to learn characteristics of the data. This is very helpful in

determining suitable learning approaches to modeling the data. To characterize the nuclear detection data, Principal Components Analysis (PCA) and Variance Analysis were performed to confirm that the data could be accurately modeled in a lower-dimensional space. For the dataset, promising results were found: it was determined that 90% of the variance was accounted for using only about 10% of the attributes. Following this, a clustering analysis was performed to determine if certain isotopes distinguish themselves from other isotopes or background "naturally" or not. Based on these experiments, it was concluded that "isotopes" and "background" could not reasonably be determined by clustering. This led to the decision to classify the isotopes individually, with the algorithms determining if a sample is a specific isotope or not (i.e. classified as "absent" or "present").

These results led to research done by Nelson and Pottenger [8]. In this work, four isotopes were focused on: Ga67, In111, I131, and Tc99m. This research focused on two classification algorithms: Naïve Bayes and Higher-Order Naïve Bayes (HONB) in order to demonstrate that Higher-Order information can be extremely useful in the arena of nuclear detection. The HONB and Naïve Bayes algorithms modeled various samples of training data. Overall, results showed that HONB was usually able to outperform Naïve Bayes. These results were consistent in all three of the metrics (Accuracy, Weighted Macro Average, and Un-weighted Macro Average). This prior work thus demonstrated that Higher Order Learning techniques, and in particular HONB, can be useful in the arena of nuclear detection.

III. Approach

This project focuses on classifying topics learned using the Latent Dirichlet Allocation (LDA) and Higher-Order Latent Dirichlet Allocation (HO-LDA) algorithms on a nuclear detection numeric dataset. HO-LDA will be described in the next section. LDA and HO-LDA learn topics on the nuclear detection dataset to obtain a topic decomposition of instances. These learned topics are then used as features in a traditional supervised classification algorithm. In essence, the LDA or HO-LDA topic assignments are used as features in supervised learning algorithms that predict the class (isotope), treating LDA or HO-LDA as feature space transforms. Two different approaches for feature creation were explored. Using Topic Modeling on numeric nuclear detection data is cutting edge, as to our knowledge this has never been done before on a nuclear detection data set.

A. Higher-Order Latent Dirichlet Allocation

In this section, we present a novel approach to topic modeling based on the Higher Order Learning framework called Higher-Order Latent Dirichlet Allocation (HO-LDA). As noted in the Background and Prior Work section above, traditional machine learning methods only consider relationships between feature values within individual data instances while disregarding the dependencies that link features across instances. In [5], a general approach to supervised learning has been developed by leveraging higher-order dependencies between features. Unlike approaches that assume data instances are independent, this framework leverages relations between feature values across different instances. Additionally, this framework can be generalized using a novel data-driven space transformation that allows any classifier operating in vector spaces to take advantage of these higher-order relations. The utility of this transform has been established in algorithms including Higher-order Naïve Bayes, Higher-order Support Vector Machines, etc.

The objective of this aspect of the proposed effort is to incorporate higher order information into the framework of LDA. We modified the Gibbs-sampling formula of LDA by replacing feature frequencies in topics with their higher order frequencies. In other words, in equation (1) we replaced these counts with higher path counts, $c_{i,j}$, for the feature *w* in topic *j*. $c_{i,j}$ is computed as follows, assuming that the input to this algorithm is a set of nuclear detection instances, each being labeled with a topic index as in [10]:

1. Partition each instance into sets of entities E_1, E_2, …, E_k according to the topics that are assigned.
2. Each topic *j* now has a corresponding set of partial communications. The higher order path counts $c_{i,j}$ are computed exactly the way they are computed for Higher Order Naïve Bayes [11], the classes corresponding to topics.

This approach holds promise to distinguish more precise topics that leverage correlation between topics without sacrificing the mathematical simplicity of LDA in favor of more complex algorithms that model topic correlation directly.

B. The Data

The data for this project was taken from a hand-held CZT (Cadmium Zinc Telluride)-based radiation detector, called the Interceptor™. The Interceptor™ is a Thermo Scientific handheld Spectroscopic Personal Radiation Detector. The dataset obtained for this project includes 302 gamma-ray spectrum files. There were instances for 17 isotopes as well as background instances included in the 302 instances. Each of these gamma-ray spectrum files contains one spectrum, and there are 1024 numeric channels per spectrum, with high dimensional space. Each channel contains an integer count of photon interaction events which were recorded within a preset detection interval, usually 60 seconds long for a hand-held device. The spectrum covered an energy range from 0 to approximately 1.5MeV.

C. Two Approaches To Feature Creation

Two approaches were applied to feature creation from the nuclear detection dataset. The methodology was nearly identical for the two methods, but the treatment of the dataset varied. The first approach to feature creation treated the nuclear detection dataset as multinomial, and is referred to in what follows as Multinomial Feature Creation. This is the common approach to topic modeling, which has been primarily based on textual data as input. The second method treated the highest number of the individual channel readings as the number of attributes, and is referred to below as Maximum Channel Value. Both approaches compared topics learned using HO-LDA with traditional LDA using various standard classification metrics.

HO-LDA and LDA in the Multinomial Feature Creation approach treat the nuclear detection data as multinomial. The data has 1024 channels, so the topic model input consisted of vectors of 1024 attributes. Topics were learned using HO-LDA or standard LDA. These learned topics were input into a traditional supervised classification algorithm, including a decision tree learning algorithm from the WEKA Workbench, J48, and Naïve Bayes. Ten trials were performed for a range of number of topics and sample size for both HO-LDA and LDA, and these results were then compared using the standard metrics of Accuracy, Precision, Recall, and F-Measure to determine statistical significance.

The Multinomial Feature Creation approach first used the full dataset with the full attribute set (1024 attributes). It was determined after obtaining some results that the attribute set could be pruned prior to learning topics. Therefore, results are mainly presented based on the use of attribute subset selection to prune the channels before learning the topic model. Again using the WEKA Workbench, the number of attributes was pruned from 1024 to only 39. This approach was performed on both the entire dataset as well as samples of training data (20%, 25%, 33%, and 50%). The number of topics chosen to be used in LDA and HO-LDA were 5, 10, 20, and 50. In addition, some results are presented at the individual class level.

In the second feature creation approach (Maximum Channel Value), the nuclear detection dataset is directly read into the topic model with the highest actual isotope channel reading counted as the number of attributes. This was done using the full set of 1024 channels. The LDA or HO-LDA results based on this input are then classified using the decision tree classifier J48. Thirty trials were performed for each sample size and topic number, and the HO-LDA results were compared to the LDA results using the standard metrics of Accuracy, Precision, Recall, and F-Measure using a t-test for statistical significance. The entire dataset was examined for this approach, as well as samples of the data (20%, 25%, 33%, and 50%). The number of topics chosen for LDA and HO-LDA were 5, 10 and 20, 50, and 100.

IV. The Results

In the Results section, results from both the Multinomial Feature Creation approach and the Maximum Channel Value approach are presented. Using Multinomial Feature Creation, HO-LDA generally performed similarly to LDA, outperforming LDA in the 100% case for the 5 and 100 topic models. Similar results were obtained for experiments with various training sample sizes. Although not conclusive, the results from Multinomial Feature Creation do nonetheless indicate that Topic Modeling can be applied to the arena of nuclear detection. In the case of the Maximum Channel Value approach, HO-LDA consistently outperformed LDA. These results are reported in subsection B following.

A. Multinomial Feature Creation

The Multinomial Feature Creation approach treated the nuclear detection data as multinomial data, with a 1024-dimension feature space. First, this approach tested the full feature space, and then a feature space reduction was performed, paring down the number of attributes to just 39. Results from both approaches are presented below.

The full feature space was examined using thirty trials to compare HO-LDA with LDA using 5, 10, 20, 50, and 100 topics. This was first performed using the full dataset. For 5 and 100 topics, HO-LDA outperformed LDA in all four metrics used (Accuracy, Precision, Recall, and F-Measure) with the single exception of Precision for 100 topics using the Naïve Bayes classifier. See Table 1 below for the Accuracy results for the J48 classifier and Table 2 for the Naïve Bayes classifier results. All four metrics were similar when using J48, and were mostly the same when using Naïve Bayes. Next, a 25% training sample size was examined using both the J48 and Naïve Bayes classifiers for 5 topics. Neither HO-LDA nor LDA were statistically significantly better in this case.

TABLE 1. ACCURACY T-TEST RESULTS – HO-LDA VS LDA WITH J48 CLASSIFIER FOR MULTINOMIAL FEATURE CREATION WITH 1024 ATTRIBUTES

#Topics	Accuracy HO-LDA Avg.	Accuracy HO-LDA Standard Deviation	P VALUE	Accuracy LDA Avg.	Accuracy LDA Standard Deviation
5	0.714	0.020	0	0.695	0.018
10	0.772	0.013	0	0.791	0.018
20	0.754	0.014	0	0.785	0.015
50	0.761	0.016	0	0.778	0.016
100	0.820	0.010	0	0.755	0.017

TABLE 2. ACCURACY T-TEST RESULTS - HO-LDA VS LDA WITH NAÏVE BAYES CLASSIFIER FOR MULTINOMIAL FEATURE CREATION WITH 1024 ATTRIBUTES

#Topics	Accuracy HO-LDA Avg.	Accuracy HO-LDA Standard Deviation	P VALUE	Accuracy LDA Avg.	Accuracy LDA Standard Deviation
5	0.686	0.010	0	0.630	0.009
10	0.793	0.010	0	0.802	0.008
20	0.840	0.008	0	0.866	0.006
50	0.818	0.010	0	0.840	0.008
100	0.821	0.006	0.016	0.815	0.011

TABLE 3. ACCURACY T-TEST RESULTS - HO-LDA VS LDA WITH J48 CLASSIFIER FOR MULTINOMIAL FEATURE CREATION WITH 39 ATTRIBUTES

#Topics	Accuracy HO-LDA Avg.	Accuracy HO-LDA Standard Deviation	P VALUE	Accuracy LDA Avg.	Accuracy LDA Standard Deviation
5	0.629	0.016	0	0.706	0.020
10	0.683	0.014	0	0.670	0.010
20	0.690	0.007	0	0.700	0.020
50	0.679	0.015	0	0.638	0.011

TABLE 4. ACCURACY T-TEST RESULTS - HO-LDA VS LDA WITH NAÏVE BAYES CLASSIFIER FOR MULTINOMIAL FEATURE CREATION WITH 39 ATTRIBUTES

#Topics	Accuracy HO-LDA Avg.	Accuracy HO-LDA Standard Deviation	P VALUE	Accuracy LDA Avg.	Accuracy LDA Standard Deviation
5	0.607	0.012	0	0.748	0.007
10	0.765	0.007	0.013	0.760	0.006
20	0.786	0.010	0	0.763	0.012
50	0.756	0.009	0	0.769	0.011

The next set of experiments performed used the pruned attribute size of 39 attributes versus 1024. Ten trials were performed for each experiment to determine if HO-LDA produced features that performed statistically significantly better than LDA. First this approach was applied on the full dataset using 5, 10, 20, and 50 topics, with classifiers J48 and Naïve Bayes. Using the J48 classifier, HO-LDA outperformed LDA with 10 and 50 topics using all four metrics. Results were fairly similar across the metrics. See Table 3 for the Accuracy results. Results using Naïve Bayes showed that while the averages for all four metrics were again similar, the statistical significance varied. Accuracy is shown in Table 4, which was also similar to Recall. However, for Precision, HO-LDA was not statistically significantly better than LDA. In fact, LDA statistically outperformed HO-LDA for 5, 10, and 50 topics. For F-Measure, HO-LDA outperformed LDA statistically significantly with 20 topics, while LDA statistically significantly outperformed HO-LDA for 5 and 50 topics.

Next, various training sample sizes were examined for the reduced attribute nuclear detection dataset. In particular, 20%, 25%, 33%, and 50% training sample sizes were examined for 5, 10, 20, and 50 topics using both J48 and Naïve Bayes.

TABLE 5. PRECISION T-TEST RESULTS - HO-LDA VS LDA WITH J48 CLASSIFIER FOR MULTINOMIAL FEATURE CREATION WITH 39 ATTRIBUTES

#Topics	Precision HO-LDA Avg.	Precision HO-LDA Standard Deviation	P VALUE	Precision LDA Avg.	Precision LDA Standard Deviation
5	0.630	0.015	0	0.705	0.021
10	0.687	0.020	0	0.667	0.012
20	0.694	0.012	0.669	0.696	0.025
50	0.678	0.015	0	0.637	0.011

classifiers. For the 20% sample size using the J48 classifier, LDA outperformed HO-LDA for 50 topics with the Accuracy and F-Measure metrics. For the rest of the metrics and across all numbers of topics, neither approach performed statistically significantly better. For the 25% training sample size, neither HO-LDA or LDA performed statistically significantly better for both J48 and Naïve Bayes. For the 33% sample size, HO-LDA performed statistically significantly better than LDA with 10 topics as measured by both the Accuracy and Recall metrics. When using the Naïve Bayes classifier, LDA statistically significantly outperformed HO-LDA with 10 and 20 topics using the Precision and F-Measure metrics. For the 50% sample size, HO-LDA statistically significantly performed better than LDA with 10 and 50 topics across all four metrics. LDA statistically significantly performed better than HO-LDA with 5 topics for all metrics, and with 20 topics for Accuracy and Recall. See Table 5 for the Precision results. For the Naïve Bayes classifier, LDA statistically significantly outperformed HO-LDA with 5 topics across all four metrics.

The next set of experiments employed the pruned attribute size of 39 to examine classification at the individual class level. This was performed for the 25% sample size with 5 and 10 topics using both J48 and Naïve Bayes.

For 5 topics, J48, HO-LDA was statistically significantly better than LDA for the isotopes Ra226 and Tl201 for Accuracy, Ra226 and Xe133 for Precision, Ra221 and Tl201 for Recall, and Ir192, In111, and Na22 for F-Measure. For 5 topics, J48, 5 topics, LDA was statistically significantly better than HO-LDA for Ga67, I123, I131, In111, Na22, and Background with Accuracy, for Tc99m, Ga67, I123, I131, In111, Na22, and Tl201 with Precision, for Ga67, I123, I131, In111, Na22, and Background for Recall, and for I131, In111, Na22, Tc99m, Ga67, and I123 for F-Measure. For 5 topics, Naïve Bayes, HO-LDA statistically significantly outperformed LDA for Th232 with Accuracy, Ra226 and Th232 for Precision, Th232 for Recall, and I125 and Th232 for F-Measure. For 5 topics, Naïve Bayes, LDA statistically significantly outperformed HO-LDA for Ra226, U235, Ga67, I123, I131, and Na22 for Accuracy, for Ba133, Ga67, I123, I131, In111, Na22, and U235 for Precision, for Ga67, I123, I131, Na22, Ra226, and U235 for Recall, and for Ga67, I123, I131, In111, Na22, Ra226, and U235 for F-Measure.

For 10 topics, J48, HO-LDA performed statistically significantly better for Ba133, I131, and Xe133 for Accuracy, for Ba133, Th232, Tl201, and Xe133 for Precision, for Ba133, I131, Xe133 for Recall, and for Ba133, I131, Th232, Tl201, and Xe133 for F-Measure.

TABLE 6. ACCURACY T-TEST RESULTS HO-LDA VS. LDA WITH J48 CLASSIFIER FOR MAXIMUM CHANNEL FEATURE CREATION.

#Topics	Accuracy HO-LDA Avg.	Accuracy HO-LDA Standard Deviation	P VALUE	Accuracy LDA Avg.	Accuracy LDA Standard Deviation
5	0.675	0.014	0	0.610	0.012
10	0.688	0.015	0	0.641	0.014
20	0.654	0.015	0.021	0.644	0.017
50	0.665	0.014	0	0.605	0.015
100	0.565	0.015	0	0.529	0.015

For 10 topics using J48, LDA statistically significantly outperforms HO-LDA for Ga67, In111, and Na22 for Accuracy, Ga67, In111, and Na22 for Recall, and for Ir192 for F-Measure. For 10 topics using Naïve Bayes, HO-LDA statistically significantly performs better than LDA for Ga67, I123, Th232, and U235 for Accuracy, for I123, Na22, Th232, and U235 for Precision, for Ga67, I123, Th232, and U235 for Recall, and for Ga67, I123, Na22, Th232, and U235 for F-Measure. For 10 topics, Naïve Bayes, LDA statistically significantly performs better than HO-LDA for Ba133 with Accuracy, for I125, I131, Ra226, and Background for Precision, for Ba133 for Recall, and for Ba133, I125, I131, Ra226, and Background.

Although these per-class results are somewhat tedious to review, they generally confirm that HO-LDA and LDA perform similarly using the Multinomial Feature Creation approach.

B. Maximum Channel Value Feature Creation

As before, the purpose of the Maximum Value Feature Creation approach was to determine if HO-LDA techniques perform better than LDA on the nuclear detection dataset. These results will give insight into the utility of HO-LDA on datasets with small number of examples.

First, trials were performed on the full dataset with 5, 10, 20, 50, and 100 topics (see Table 6). Thirty trials were performed for each experiment (i.e., thirty trials were run for 5 topics HO-LDA, and thirty J48 classification trials were run for 5 topics LDA) to determine statistical significance using standard 5 fold cross validation. The Accuracy metric was the one used for these experiments. In these experiments, HO-LDA statistically significantly outperformed LDA in all cases.

Next, experiments were performed using randomized stratified training samples of nuclear detection data, and then the remaining results were folded in to the LDA or HO-LDA model in order to perform classification on all 302 samples. Training sizes included 20%, 25%, 33%, and 50%, and the number of topics included were 5, 10, and 20. Again, thirty experiments were performed for each test. Metrics used included Accuracy, Precision, Recall, and F-Measure. With all four training sample sizes, HO-LDA statistically significantly outperformed LDA in all cases with all four metrics, with the exception of the 50% training size, 5 topics, with the Accuracy and Recall metrics. These results illustrate that when using the Maximum Channel Feature Creation approach, HO-LDA statistically significantly outperforms LDA, especially on small samples of training data.

V. CONCLUSIONS

In this paper, a novel approach to topic modeling based on the Higher Order Learning framework is introduced. Higher-Order Latent Dirichlet Allocation is presented, along with its application to an important issue in homeland security, nuclear detection. In addition, this work strives to improve topic models in the real time environment of online learning. Using topic modeling on numeric nuclear detection data is also cutting edge, as to our knowledge this has never been done before. Two methods of feature creation were evaluated, including Multinomial Feature Creation and Maximum Channel Value Feature Creation. The Multinomial Feature Creation Approach demonstrates that although topic modeling can be usefully applied to the arena of nuclear detection, HO-LDA performed similarly to LDA. In contrast, results for Maximum Channel Value Feature Creation illustrate that when using this approach, HO-LDA consistently statistically significantly outperforms LDA, especially on small samples of training data. HO-LDA outperformed LDA for the entire dataset for the number of topics we chose, as well as most of the training sample sizes selected. This is an important first milestone in the application of Higher Order Learning in the domain of topic modeling.

ACKNOWLEDGEMENTS

The authors acknowledge the contributions of Christopher D. Janneck and Jason Perry for their previous work and assistance on this project. This research was supported by the U.S. Department of Homeland Security under Grant Number DHS-28DN077ARI012-04. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DHS.

We are also grateful for the help of co-workers, family members, and friends. Co-author W. M. Pottenger also gratefully acknowledges the continuing help of his Lord and Savior, Yeshua the Messiah (Jesus the Christ) in his life and work.

REFERENCES

[1] Carpenter, Tamra, Cheng, Jerry, Roberts, Fred, and Xie Minge. (2009) "Sensor Management Problems of Nuclear Detection." , Unpublished manuscript.

[2] Ganiz, Murat Can, George, Cibin, and Pottenger, William M. (2011) "Higher Order Naïve Bayes: A Novel Non-IID Approach to Text Classification", IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 7, pp. 1022-1034, July, doi:10.1109/TKDE.2010.160.

[3] Perry, Jason. (2009) "Clustering and Machine Learning for Gamma Ray Spectroscopy.", Unpublished manuscript.

[4] Thermo Scientific. User Manual Interceptor™ Spectroscopic Personal Radiation Detector. Oct 2011.

[5] Ganiz, M. C., Lytkin, N. I. and Pottenger, W. M. (2009) Leveraging Higher Order Dependencies Between Features for Text Classification. In the *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (ECML PKDD). Bled, Slovenia, September.

[6] Blei, David M. "Introduction to Probabilistic Topic Models." *Communications of the ACM*, to appear.

[7] Pottenger, William M., Kantor, Paul, Li, Shenzi, Kolipaka, Kashyap, and Pandya, Chirag. "Final Report on Entity Resolution System" U.S. Department of Justice, National Institute of Justice, Information Led Policing Research, Technology Development, Testing, and Evaluation.

[8] Nelson, Christie and Pottenger, William M. "Nuclear Detection Using Higher Order Learning." In the proceedings of the *IEEE International Conference on Technologies for Homeland Security*. Boston, MA: Nov, 2011.

[9] Blei, D., Ng, A., and Jordan, M. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993-1022. 2003

[10] Aitchison, J. The statistical analysis of compositional data. *Journal of the Royal Statistical Society*, Series B, 44(2), 139-177. 1982

[11] Ganiz, M., and Pottenger, W.M. A Novel Bayesian Classifier for Sparse Data. *IEEE Transactions of Knowledge and Data Engineering (TKDE)*, 2010.