DOMAIN ADVERSARIAL NEURAL NETWORKS APPLIED TO LASER-INDUCED BREAKDOWN SPECTROSCOPY.

T. Mullen[1], M. Parente[1], and M. D. Dyar[2], [1] Department of Electrical & Computer Engineering, University of Massachusetts, Amherst MA 01003 thmullen@umass.edu, [2]Department of Astronomy, Mount Holyoke College, 50 College St., South Hadley, MA, 01075.
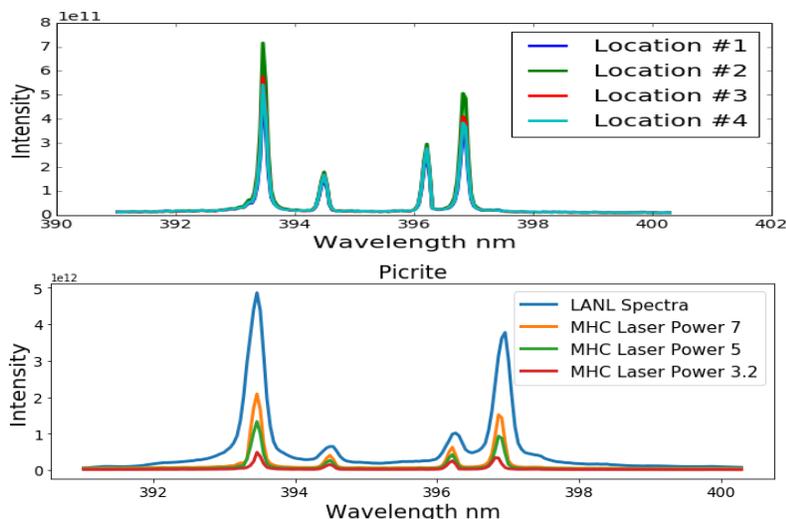
Introduction: Laser-Induced Breakdown Spectroscopy (LIBS) is a technique used to quantify elemental abundances. Variabilities in the signal that stem from sources other than elemental composition [1, 2] pose major challenges for LIBS quantification. These include shot-to-shot variations and sample heterogeneity, seen as signal variations in different locations on the same sample (Figure 1 top). Although such variations pose challenges to many muyltivariate analysis techniques used for quantification of LIBS results, deep neural networks built from linear functions passed through non-linear activation layers show promise. Consisting of many more tunable parameters than other regression techniques, neural networks have the capability to learn complicated transformations that can handle the amount of variability seen in LIBS. Our team has previously shown that a single four-layer neural network can regress from the full spectra to all nine major oxides [3], learning from all spectral shots rather than just averages from each sample or location. When trained on mean squared error, such a network attains similar performance to the other traditional regression techniques such as partial least squares (PLS) regression.

Variations in laser power on target can also cause spectral variability. Spectra of the same sample taken at different laser powers have easily identifiable variations (Figure 1b) caused by the different populations of emission lines that are present at varying plasma temperatures. Many other characteristics of an instrument, especially including detector sensitivity, can cause differences in resulting spectra (Figure 1 bottom) that complicate efforts to use a model built on one instrument to interpret data acquired on a different one.

Clegg et al. [4] dealt with the issue of transferring from Los Alamos National Laboratory (LANL) spectra to Martian spectra by creating a ratio to apply to the LANL spectra to make them more like the Martian spectra. Using six LANL and Martian spectra from the same samples, they created six ratios. They averaged the six to create one ratio to apply to whole LANL dataset. This allows a model trained mostly on terrestrial data to be used on Martian spectra. To test an analogous problem with more labeled data, Mullen et al. [3] showed that training a neural network on 1329 samples from the Mount Holyoke College (MHC) ChemLIBS instrument and testing on 163 LANL samples had poor performance. To transfer from different domains, they fine-tuned an MHC-trained network using 63 LANL samples. This network performed better, but still did not achieve the same results as training and testing on LANL samples.

The Experiment: To mirror the problem of adapting a regressor trained on terrestrial data to Martian data, we set up an experiment to adapt a regressor trained on MHC ChemLIBS data to LANL data acquired using a ChemCam-like instrument [4]. We trained regressors on spectra acquired using two laser powers (2.8 and 3.8 mJ) as well as 11 samples from LANL (95A, ca. 10 mJ [5]). The MHC-ChemLIBS dataset is made up of spectra with five shots each taken at six locations on each sample. There are 50 shots taken at each of five locations in the LANL dataset.

For labeled data in our model, we used the MHC data plus the 11 samples from Los Alamos; the remaining Los Alamos data were unlabeled so no labels were utilized to train the model. We did use the remaning 250 samples in the LANL

Figure 1: LIBS data from sample PICRITE acquired at Mount Holyoke on the ChemLIBS instrument. Top: Although this pellet is composed of finely-ground powder, spectral response varies among different analysis locations. Bottom: Spectra of sample PICRITE taken by instruments at LANL and MHC, the latter using different laser powers (7% = 3.8 mJ, 5% = 2.8 mJ, and 3.2% = 1.9 mJ).

dataset to tune hyperparameters. There is also a holdout test set of LANL data that will be used in the future for additional testing.

Partial Least Squares and Ratio Domain Adaptation: To approximate methods currently being used by the ChemCam team on *Mars Science Laboratory*, we trained a PLS model on the MHC dataset. Before training the model we applied a ratio to MHC. To create the ratio we followed the work of Clegg et al. [4] explained above. We chose 12 samples that were analyzed at both LANL and MHC, the latter using two laser powers. We found the ratio between spectra of the same sample from different laser powers and then averaged the ratios resulting from the overlapping samples. That ratio was applied to all MHC spectra. Using the corrected spectra in the labeled dataset, a PLS model was trained. We varied the number of components used from 0 to 20, used a five-fold cross-validation scheme on the labeled data, and tested the model on the LANL validation set. The best model was chosen for each element using cross validation and LANL validation errors. LANL validation errors are listed in **Table 1**. It is important to note that these models were all trained only with averaged spectra. For MHC, the spectra were averages over locations and for LANL, they were averages over the whole sample.

Element	RMSE
Al_2O_3	4.22
CaO	2.35
Fe_2O_3	4.60
K_2O	1.07
MgO	2.93
Na_2O	1.00
Si_2O	7.31
Ti_2O	0.75

Table 1: Root mean squared error resulting from the PLS model.

Domain Adversarial Neural Networks: As noted by Ganin et al. [6], domain adversarial networks choose features that are optimal for regression, but not discriminative of which dataset they came from. They seek to preserve information that helps with the regression problem while ignoring the signal that comes from the instrument and instrument settings. To build these features from the LIBS spectra, we use a neural network with an adversarial loss function. The network can be thought of as three different parts. The first is a feature extractor that transforms the spectra to a feature vector using a three-layer neural network. The next is a regression network that takes the features and predicts the elemental abundances. The last is a domain classifier, which uses the features to predict which instrument the original spectra came from. Both the regression and domain classifier networks are one single output layer.

Neural networks are trained by backpropagating the gradient of the loss function with respect to the weights and updating the weights in the direction of gradient descent. Instead of backpropagating the gradient of the domain classifier to the feature extractor and using gradient descent, this system changes the weights in the direction of gradient ascent. By reversing the gradient, the system is forced to change the weights in the feature extractor to make the features more difficult to classify. The reversal is just applied to the gradient coming from the domain classifier loss. Note the gradient updating the domain classifier layer use gradient descent; the only update in the ascent direction is the gradient domain classifier loss with respect to the feature extractor weights. This optimizes the domain classifier for the features it is given, as well as the regressor.

The domain classifier can use unlabeled data to make the features less informative about which instrument created them. This is a significant advantage over other regression techniques because the domain classifier does not allow the regressor to overfit to the labeled dataset. By constantly changing the features based on the input from the unlabeled spectra, a regressor that works for both datasets can be produced.

Results: The domain adversarial network attained better performance than PLS on Fe_2O_3, MgO on the validation set and was close on the remaining elements (**Table 2**). RMSE's in the neural network were taken over all shots and locations not just the average spectra as in the PLS model. The neural network did not require any transformation of the spectra and regressed to all eight elements at once. The domain adversarial neural network was also able to utilize the unlabeled spectra making it a well-suited tool for Mars where there will be little labeled data.

Element	RMSE
Al_2O_3	4.61
CaO	3.37
Fe_2O_3	3.96
K_2O	1.86
MgO	1.96
Na_2O	1.86
Si_2O	7.95
Ti_2O	1.35

Table 2: Root mean squared error resulting from the domain adversarial NN model.

References: [1] Clegg S. M. et al. (2009) *Spectrochim. Acta B*, 64, 79-88. [2] Anzano J. M.et al. (2006). Analytica Chimica Acta, 575, 230-235. [3] Mullen T. et al. (2017) *AGU*. [4] Clegg S. M. et al. (2017) *Spectrochim. Acta B*, 129, 64-85. [5] Wiens R. C. et al. (2013) Spectro. Acta B, 82, 1-27. [6] Ganin Y. et al. (2016) *Journal of Machine Learning Research,* 17, 1-35.