Machine Learning in a Quality-Managed RF Measurement Workflow*

Aric W. Sanders[1,5], John M. Bass[2], Arpita Bhutani[3], MaryAnn S. Ho[4], James C. Booth[5]

[1]University of Colorado, Department of Physics, Boulder, CO, [2]Rose-Hulman Institute of Technology, Terre Haute, Indiana, [3]Peak to Peak Charter School, Lafayette, Colorado, [4]Fairview High School, Boulder, Colorado, [5]National Institute of Standards and Technology, Boulder, Colorado

***Abstract* — Advances in artificial intelligence, or more specifically machine-learning, have made it possible for computers to recognize patterns as well as or better than, humans. The process of quality management in radio-frequency measurements is an arduous one that requires a skilled human to interpret complex data sets and frequently leads to long diagnostic routines when a measurement fails a quality check, or an operator makes a mistake. We demonstrate the application of machine-learning classifiers to improve a common task in a quality managed measurement workflow. Specifically, we demonstrate that a machine-learning classifier can predict the device attached to a measurement apparatus and additionally predict if one of the connections has been improperly torqued, furthermore predicting which connection is loosened.**

***Index Terms* — Artificial intelligence, machine-learning, quality management, rf metrology, verification standards.**

I. Introduction

Recent progress in the quality and ease of implementing machine learning in computer science has made the technology ubiquitous in our daily lives. From cell phones to smart speakers, this technology makes it possible for computers to recognize patterns as well as or better than, humans in some cases. Simultaneously, the complexity of radio-frequency measurements in communications and elsewhere has increased in complexity as measurements are extended to higher frequencies, larger number of ports, and non-linear devices. The process of insuring quality measurements in these complex situations largely falls upon well trained humans that make judgments based on quality management protocols established over many years or decades [1]. While these quality protocols have proven to be reliable for measurements with few process variables, they are costly in terms of time and labor when a system or device fails, normally signaling a non-compliance event that must either be rectified by re-measurement or become the subject of an extensive review. This process could be revolutionized through the implementation of the disruptive technologies in machine learning. In the future, artificial intelligence agents could be trained to recognize common faults in complex systems, informing humans of the most likely cause of a non-conformance event and even suggesting the best course of action. However, before this future is realized, the most promising technologies must be tested and proven to be capable of the performing tasks that humans currently do.

Here we demonstrate a methodology to train machine-learning classifiers to perform the task of identifying a connected device under test (DUT) and to determine whether that device has been properly torqued at each of several points in an adapter chain. We test eight popular machine-learning classifiers, first training them to recognize the raw, uncorrected scattering parameters of a system of adapters and calibration devices. We further extend this training to a set of states or labels where specific adapter connections have been intentionally loosened to an uncontrolled torque. Next, we assess the trained machine-learning classifiers against two training accuracy data sets and one test accuracy to determine their accuracy in determining the DUT, the state of the connections to that DUT and the precise location of loosened connection.

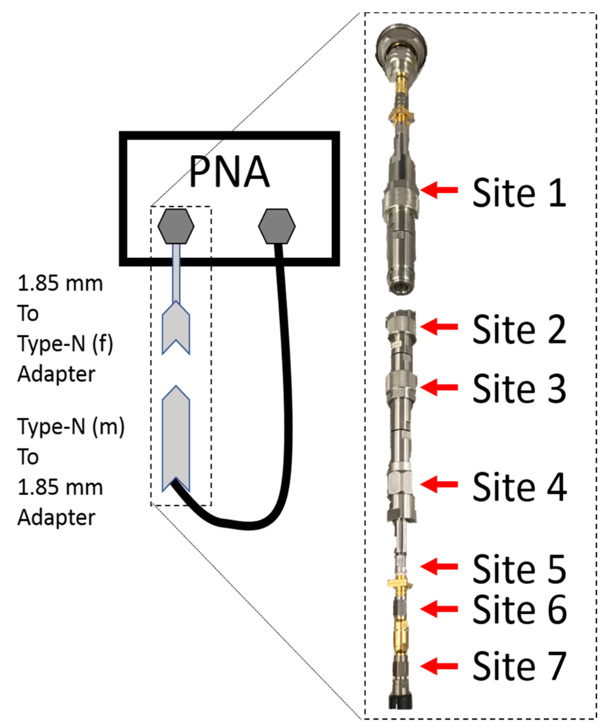

Fig. 1. Experimental setup for training measurements with connection sites labelled. The precision network analyzer (PNA) is adpated from its 1.85 mm port connector to the device under test's type-N connector.

978-1-7281-0506-2/19/$31.00 ©2019 IEEE

II. Experimental Setup

In order to establish the statistics of different DUTs in properly and improperly torqued connecter states, a standard type-N Short–Open–Load-Thru (SOLT) insertable calibration kit (Agilent** 85054B) was measured on an Agilent E8631A precision network analyzer (PNA). The measurements were of standard two port scattering parameters, carried out using an intermediate frequency of 10 Hz, a nominal power of 10 µW (-20 dBm), with no averaging. The frequency of the measurements spanned from 100 MHz to 18 GHz in frequency increments of 895 MHz for a total of 201 points. Fig. 1 shows the standard setup that consists of a cable and adapters that step from the 1.85 mm connector of the PNA ports to the type-N connecter of the calibration devices. Seven sites separated from the instrument ports were chosen to sequentially loosen, measure, and then torque to specification (1.4 Nm for type N, 0.9 Nm for all other connectors) and measure again. The inset in Fig. 1 shows an exploded view of the insertable test ports with each site labelled. Data in both the properly torqued and loosened conditions was taken without a correction applied.

III. Classifier Training

In order to train the classifiers of interest, a large amount of data representing the patterns of interest should be used. The amount of data to accurately predict the state of the measurement is typically unrealistic to acquire on a laboratory time scale, hence we increase the amount of data used in training by determining the statistics of each device and state of connections and then use these statistics to randomly generate more robust training data. We do this by first acquiring a pre-training set that has at least three repeated measurements of each DUT (S, O, L, and T) and connector state for a total of 32 states or labels. Each repeated measurement is a single acquisition in a connect / disconnect cycle, where the site of interest is loosened by approximately 30 – 50 degrees from the properly torqued state and data is collected, after which the experiment is returned to the properly torqued state. After each loosen cycle, data for the DUT in the properly torqued state is recorded. This leads to three measurements of the loosened state and between 10-25 measurements of the properly torqued state for each DUT, for a total of 173 measurements of the 32 labels or states.

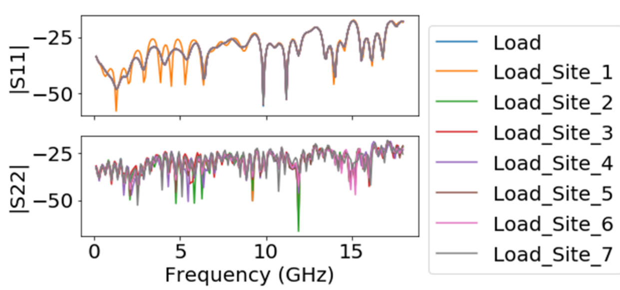

Fig. 2. Measurement of the amplitudes of the reflection coefficients in dB for 8 states of the load device.

In Fig. 2, the magnitude of the reflection coefficients in logarithmic units for 8 of the pre-train data measurements for a broadband load are shown. It is interesting to note that the measurements show only small differences for the loosened cases. The mean and standard deviations of these measurements are then calculated for each real and imaginary component of the scattering parameter at each frequency. A set of 1000 training sets for each state is then generated by choosing a random number from a Gaussian distribution with the same mean and standard deviation as the pre-training data. A small value, 1.0 $*10^{-9}$ was added to the standard deviation of the Gaussian to insure no points would have zero variation. The training set is then vectorized or flattened to a 1-d array containing the real and imaginary parts of each scattering parameter so that each training set is 1608 points.

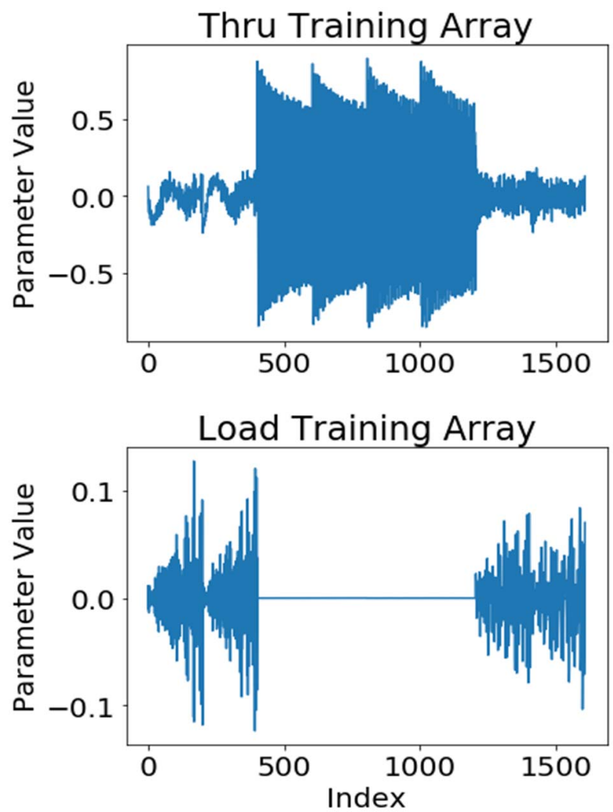

Fig. 3. A typical vectorized training set for the thru and load standards. Each training set has the real and imaginary components for each of four scattering parameters.

Fig. 3. shows a typical vectorized training set for the torqued thru and load. It should be noted that because the training of a classifier is a probabilistic process, classifiers of the same type will have slightly different prediction results for the same pre-train data. The classifiers and the training process

TABLE I

CLASSIFIER RESULTS

		Classification Method							
	Metric	Quadratic Discriminant	Bernoulli Naïve Bayes	Decision Tree	Random Forest	Ridge	Stochastic Gradient Descent	Passive Aggressive	Perceptron
Training Accuracy (Training Data)	State Accuracy	100.0%	100.0%	100.0%	100.0%	100.0%	78.1%	93.8%	96.9%
	Device Accuracy	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	96.9%
	Connection Accuracy	100.0%	100.0%	100.0%	100.0%	100.0%	96.9%	100.0%	96.9%
Training Accuracy (Pre-train Measurements)	State Accuracy	64.7%	100.0%	99.4%	100.0%	99.4%	86.7%	95.4%	83.8%
	Device Accuracy	100.0%	100.0%	100.0%	100.0%	100.0%	95.4%	99.4%	96.1%
	Connection Accuracy	67.1%	100.0%	99.4%	100.0%	99.4%	90.2%	98.8%	84.4%
Test Accuracy (Test Measurement Data)	State Accuracy	25.5%	96.1%	82.4%	74.5%	86.3%	86.3%	86.3%	76.5%
	Device Accuracy	100.0%	100.0%	100.0%	100.0%	94.1%	100.0%	100.0%	100.0%
	Connection Accuracy	64.7%	100.0%	100.0%	100.0%	94.1%	100.0%	100.0%	100.0%

were implemented in the Python programing language using the scikit-learn [2] and numpy [3] packages.

IV. ACCURACY ASSESSMENT OF CLASSIFIERS

Eight classification methods were tested, a quadratic discriminant analysis, Bernoulli naïve Bayes, decision tree, random forest, ridge, stochastic gradient descent, passive aggressive, and a perceptron all of which are covered in detail elsewhere [4]-[5]. Each of the eight classification methods were used to train a classifier and to predict the particular states or labels of three different test data sets. Fig. 4. shows the role of each data set in the process of training and testing. The classification labels consisted of the DUT name and the site of a loosened connection, for instance short and short_site_1. Theses labels or states were sub-divided into two other categories after classification, device type and the presence of any loosened connection. The first data set is a copy of the data set used to train the classifiers (32000 sets) and represents a benchmark test, or training accuracy, that determines if the classifier is suited to this type of task. The second training accuracy set are the measurements of the devices used to determine the statistics of the states, or the pre-train data (173 sets). These measurements only determine the mean and standard deviation of the training data set and are never used in training unaltered. This training accuracy assessment is meant to determine the ability of the classifier to discern the different

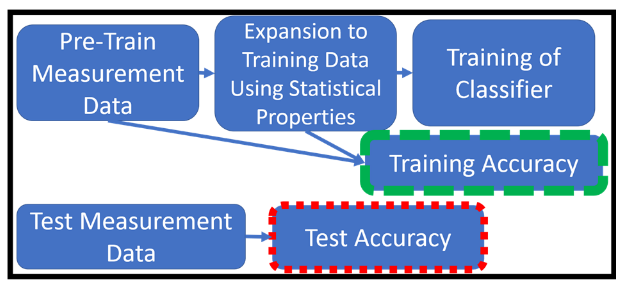

Fig. 4. Process of training and testing of the classification techniques. The training data is derived from the pre-training measurements and used to create a trained classifier. The training accuracy of the classifier is assessed by comparing the labels predicted for pre-training measurements and the training data sets. The testing of the classification technique is done with a separate set of test measurements.

states. Finally, a set of new measurements taken at a different time are used to assess the accuracy of this method for data completely independent of the training process (53 sets), which we denote as test measurement data. The results of the tests are subdivided into three categories, the total accuracy of predicting the state (DUT and any loosened location), the accuracy of predicting the DUT attached, and the accuracy of predicting if the DUT is in a properly torqued state without regard to the location of the loosened connection.

**Certain commercial equipment, instruments, or software are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

V. Results and discussion

Table I. summarizes the results of the training and test accuracy assessments. The first training accuracy set, or what the trained classifier predicts when operating on the training data set, reveals that quadratic discriminant analysis, Bernoulli naïve Bayes, decision tree, random forest, and ridge classifiers can with 100% accuracy recover the 32 states in the training data. The training accuracy statistics are calculated for all of the 32,000 training sets. This is quite promising for the prospect of using these types of classifiers and statistically similar data sets to distinguish small variations in experimental data and link them to experimental states. The other three classifiers did well overall, but sometimes devices and states that had similar characteristics, such as terminations that have no transmission, are misclassified. For instance, the perceptron classifier misclassifies several short training set as being from a load.

When applied to the measurements used to determine the statistics for the training set, or the pre-train data, we observe that five of the classifiers have training accuracy exceeding 95% in the prediction of the specific state of the 173 experimental measurements and all classifiers are capable of determining the device type with greater than 95% accuracy. This indicates the process of determining the statistics of the measurements and statistically extending the training data does not cause significant degradation in the classifier's ability to recognize the DUT, and for select classifiers there is a strong ability to determine loosened connection location based on the limited statistics of three measurements. However, this training accuracy assessment only represents how well a statistically interrelated data set is labeled by the trained classifiers; to test the prediction capability, we must use a separate set of measurements not linked to training.

The test set of 53 measurements of randomly sampled devices and loosened connections assesses the potential not only for classification of the devices and connection errors based on known experimental conditions, but also the ability to do this for future experiments. Six classifiers predicted the device connected and that there was some connection error with perfect fidelity. The Bernoulli naïve Bayes was capable of locating the connection error 96.1 % of the time, and in the event it misclassified the location of the loosened condition, it classified it as a neighboring connection. In fact, of all the tested data sets the Bernoulli naïve Bayes only misclassified two of the thru measurements loosened at site 3 as being loosened at site 2. Although these tests confirm that a machine-learning classifier has the potential to locate and inform the user of known errors, more research into the long-term stability of this type of classification is needed before being implemented into a lab quality management process.

VI. Conclusion

We have tested the ability of eight machine-learning classification techniques to determine if a loosened connection is present, what device is attached to a measurement setup, and the location of any loosened connection. Six trained classifiers were capable of accurately predicting the presence any loosened connection 100% of the time, determining the connected device 100% of the time and the determining the exact location of a loosened connection > 74% in 53 randomly sampled measurements. The best performing classifier in this set of tests was the Bernoulli naïve Bayes, which only misclassified two measurements as being from a neighboring loosened connection.

In addition to determining the test accuracy of these machine learning classification techniques, we report on two training accuracy assessments for our training protocol. The pre-training data that consisted a total of 173 measurements of 32 states, or four DUTs with eight connector configurations, were used to create a statistically similar training set of 32,000 labeled patterns. Once the machine-learning classifiers were trained with these statistically similar data, four had training accuracy identifying the presence of any loosened connection > 99% of the time, identifying the connected device 100% of the time and the location of the loosened connection > 99% of the time when used in a training accuracy assessment of the original 143 measurements. This demonstrates that the process of creating synthetic training sets does not prevent proper classification of the original measurement data.

In summary, we find that the quality management process in radio frequency measurements can be extended using techniques in machine learning to identify devices, if all connections are torqued properly, and even suggest the most likely position of any loosened connections.

Acknowledgement

The authors wish to thank Michael Schwartz of Cal Labs Solutions for insight and conversations regarding artificial intelligence in calibrations. In addition, we acknowledge the summer high-school internship program (SHIP) at NIST for support of John Bass, Arpita Bhutani and MaryAnn Ho.

References

[1] ISO/IEC 17025:2005, "General requirements for the competence of testing and calibration laboratories", International Organization for Standardization, 2nd Ed. 2005

[2] Pedregosa et al., "Scikit-learn: Machine-learning in Python" , JMLR 12, pp. 2825-2830, 2011.

[3] Travis E. Oliphant. *A guide to NumPy*, USA: Trelgol Publishing, (2006).

[4] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer Publishing, 2008

[5] S.Shlaev-Shwartz and S. Ben-David, *Understanding Machine-learning Theory and Algorithms,* Cambridge University Press, 2014