Spectral Salience

05/01/2020

Nathan Cox - [ncox4@students.towson.edu](mailto:ncox4@students.towson.edu)

Nahom Getaneh - [ngetan1@students.towson.edu](mailto:ngetan1@students.towson.edu)

James Hooper - [jhoope3@students.towson.edu](mailto:jhoope3@students.towson.edu)

Luis Ramos - [lramos6@students.towson.edu](mailto:lramos6@students.towson.edu)

Nick Taormino - [ntaorm1@students.towson.edu](mailto:ntaorm1@students.towson.edu)

Brad Wilfong - [bwilfo1@students.towson.edu](mailto:bwilfo1@students.towson.edu)

Team Leader: Dr. Lin Deng - [ldeng@towson.edu](mailto:leng@towson.edu)

Technical Director: Will Price - [cwprice@mitre.org](mailto:cwprice@mitre.org)

**Problem Title**: RF waves recognition

**Area Description**: Radio Frequency classification

**Keywords**: Burst interference, DVB-S2 protocol, Metadata, SigMF, Machine Learning, Neural Networks, Recurrent Neural Network.

**Project Description**: Our group developed a RNN model using Tensorflow to input and classify radio frequency signals. We had received all of the required technologies, our team had time to set up all the necessary software and were then ready to experiment with data. But due to the circumstances of the university’s campus closure, we were unable to proceed to capture our own data. We opted for using publicly available data from DeepSig. Then our team also decided to use a RNN in order to read in and classify the data using Tensorflow. Having previous experience with artificial neural networks, and researching various types of neural networks, we believe that the best option for classifying our data would be using this model. One of the softwares that required the use of GNU radio through the Ubuntu Linux distribution. This is the software used to send the data that we can classify.

# Executive Summary

**Title**: Spectral Salience

**Date**: 05/01/2020

**Collaborators**

|  |  |  |
| --- | --- | --- |
| **Roles** | **Names** | **Emails** |
| Technical Director | Will Price | [cwprice@mitre.org](mailto:cwprice@mitre.org) |
| Team Leader | Dr. Lin Deng | [ldeng@towson.edu](mailto:leng@towson.edu) |
| Investigator | Nathan Cox | [ncox4@students.towson.edu](mailto:ncox4@students.towson.edu) |
| Investigator | Nahom Getaneh | [ngetan1@students.towson.edu](mailto:ngetan1@students.towson.edu) |
| Investigator | James Hooper | [jhoope3@students.towson.edu](mailto:jhoope3@students.towson.edu) |
| Investigator | Luis Ramos | [lramos6@students.towson.edu](mailto:lramos6@students.towson.edu) |
| Investigator | Nick Taormino | [ntaorm1@students.towson.edu](mailto:ntaorm1@students.towson.edu) |
| Investigator | Brad Wilfong | [bwilfo1@students.towson.edu](mailto:bwilfo1@students.towson.edu) |

**Keywords**: Burst interference, DVB-S2 protocol, Metadata, SigMF, Machine Learning, Neural Networks, Recurrent Neural Network.

# Table of Contents

|  |
| --- |
| 1. Project Summary Page ……………………………. 1 |
| 2. Executive Summary Page ………………………… 2 |
| 3. Table of Contents ..…………………………………. 3 |
| 4. Introduction.…………………………………………. 4 |
| 4.1. Problem statement…………………………….... 4 |
| 4.2. Purpose statement ..………………………….… 4 |
| 4.3. Motivation ………………………...……………...4 |
| 5. Literature Review …………………………………..5 |
| 5.1 Alternate views …………………………………...6 |
| 6. Method and Procedure…………………………….7 |
| 6.1 Characterization of methods…………………...7 |
| 6.2 Procedure………………………………………..7 |
| 6.3. Schedule ………………………………………....7 |
| 6.4.Deliverables ………………………………..…….8 |
| 6.5 Limitations and Delimitations…………………..9 |
| 7. Findings…….. ………………………………….….9 |
| 7.1 Overview ………………………………….……..9 |
| 7.2 Details ………………………………….……….9 |
| 8. Issues ……………………………………………….10 |
| 9. Conclusions and Recommendations……………….11 |
| 11 References ………………………………………..12 |
| 12. Team ………………………………………….....13 |

# Introduction

**Problem Statement**

The overall issue of this project is to detect and capture RF wavelengths and be able to train a system to recognize patterns and classify and label the variation of input signals. Ideally we would use a RNN model using a Tensorflow method to achieve this goal.

**Purpose Statement**

The overall goal is to create a RNN model learning system that takes in RF waves as input, which detects and classifies them within the spectrum of RF waveforms. This system should then be able to produce an output of a labeled RF snapshot.

**Motivation**

There is an increasing demand for the use of wireless devices like mobile phones, Internet of Things (IoT), and Wi-Fi. Due to the influx of devices, the electromagnetic spectrum has become extremely crowded. The research and data that we find with the development of this project could contribute to counter security threats posed by rouge or unknown transmitters, or certain characteristics of the transmitters.

# Literature Review

Guard Band Machine Learning:

There has not been very extensive research on the data capture and labeling of RF signals. However we have found several articles and some data that have similar ideas. The investigation of indoor localization with FM radio stations investigates and evaluates the feasibility of indoor localization using broadcasting FM radio stations. The provided experimental evaluations used in this research paper provides a valid technique in the machine learning process.

Machine Learning Techniques for Cooperative Spectrum Sensing in Cognitive Radio Networks

The use of machine learning to classify cognitive radio networks is a fairly recent idea. The paper written proposes using a cooperative spectrum sensing algorithm to classify the radio network. They train using multiple classification techniques. The Css techniques are more optimized and retain a learning ability. Using the information from this paper, we could take the algorithms and ideas and apply them to our research.

Machine Learning in Adversarial RF Environments

In the RF domain, only few machine learning techniques have shown to exhibit a promising result, since the idea of RFML is fairly new. Out of those techniques, the paper analyzes five different ML techniques: Support Vector Machine (SVM), Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Non-parametric Bayesian classifier. The paper separates them into supervised and unsupervised learning and assesses the strengths and weaknesses for each of them. We are going to use this paper to determine which ML technique is best for us to use, based on the type of dataset we are working with.

Machine Learning Approach to RF Transmitter Identification

Capturing RF data and applying it to machine learning is still a fairly new concept and it can be hard to decide which machine learning concepts to use for this data. In this paper they analyse four different types, Support Vector Machines (SVM), Deep Neural Nets (DNN), Convolutional Neural Nets (CNN) and Multi-Stage Training (MST). They conducted two tests with 90% training and 10% testing for the first test and 10% training and 90% testing with the second. For this type of problem they found that for both tests the Multi-stage training (MST) was performed with the Convolutional Neural Nets (CNN) coming in close second for accuracy.

Machine Learning to Data Fusion Approach for Cooperative Spectrum Sensing

This paper covers the basics of cooperative spectrum sensing and how it improves detection performance by exploiting spatial diversity. It also compares 4 different supervised machine learning classifiers. Out of the 4 classifiers, K-nearest neighbor (KNN) and Decision Tree (DT) were found to outperform the others in how accurately they captured new frames. To come to this conclusion, a hypothesis test was conducted classifying 1000 frames and using 1000 frames to train the classifiers. This test also showed that KNN and DT classifiers had a higher, more accurate detection rate of the positive classes or spectrum holes.

Policy Based Synthesis: Data Generation and Augmentation Methods for RF Machine Learning

This research paper explores the topic of current data generation methods for RFML and the complications associated with them. More specifically, this research covers a method that would permit machine learning algorithms to better focus on salient features by using policy-based synthesis (PBS) along with existing approaches for RFML. According to their studies, this hybridized approach would enhance ML algorithms by addressing the issues with current generation methods.

**Alternate Views**

An alternate view that is present in the article “Machine Learning Approach to RF Transmitter Identification”, they mention that the best way to conduct a project like this is to create a Deep Neural Network (DNN) with Multi-Stage Training (MST). We decided to not go with this idea and instead focus more on DNN and Revolutionary Neural Network (RNN). The reason for this is that there are more libraries readily available to us for RNN than there are for the MST. While MST is great at handling high amounts of data it is relatively unknown compared to other ML methods and could cost time that we don’t have in exchange for not so much higher accuracy.

# 

# Methods and Procedures

**Characterization of Methods**

Our best approach with labeling and classifying RF signals is going with a RNN model using a TensorFlow. This suits the needs of the scope of this project because the Tensorflow libraries have the tools we need to train the ML software. In the research paper *Machine Learning Approach Transmitter Identification* they also mentioned using this model with successful results. So we hope with using the data set given and this technique for machine learning we will also be able to provide successful results.

**Procedure**

1. Gather data

* Prior data that was provided, techniques and implementations of machine learning

1. Configure hardware/software

* Create a controlled environment that emits and receives RF signals

1. Develop framework for machine learning software

* Code and develop a program that will produce statistical models of the RF spectrum

1. Train machine

* With prior data given feed the algorithm with labeled information to create a basis to compare additional RF signal data input

1. Provide summary of findings

**Schedule**

**Dates Activities**

2/17 - 2/20 Research

2/21 Project Proposal Due date

2/28 Gain familiarity with spectrum-analyzing software, and hardware.

Setup working environment (i.e. Linux Machine, data storage devices)

3/6 Run test to gather data with the Ettus B200 mini, layout our machine

learning network

3/13 Run more test and start developing a machine learning program

3/20 Spring Break

3/27 Label data and continue development in the ML program

4/3 Progress Report

4/10 Finish developing the ML program and train it

4/17 Continue training the program

4/24 Report our finding

5/1 Presentation Preparation/catchup week

5/8 INSuRe presentation

5/15 Final presentation

**Deliverables**

The proposed framework is expected to have the following features:

1. Ability to generate output waveforms as specified by the user
   1. Input source (audio stream, video stream, SDR stream, capture file)
   2. Modulation, protocol, configuration of sub-layers
   3. Signal quality (signal strength, noise floor level)
   4. Burst interference
   5. Combine waveforms
2. Documented format for metadata file associated with capture files, automatic population of metadata to document steps taken to create the waveform. For example, SigMF creates an associated meta file with JSON formatting, but would need to be extended with additional fields to describe the sub-layers of OSI Layer 1.
3. Database capable of storing waveforms and searching associated metadata
4. Application programming interfaces (APIs) to allow for generating waveforms with new protocols and configuration options
5. User interface that allows for searching capture file metadata and generating waveforms

**Limitations and Delimitations**

One of the biggest limitations was the time that we had to work on with this project, as well as not being able to physically meet and work with the required hardware.

As for delimitations, since we no longer are capable of using the hardwares to generate our own dataset, we resort to using RF dataset provided by DeepSig.

# Findings

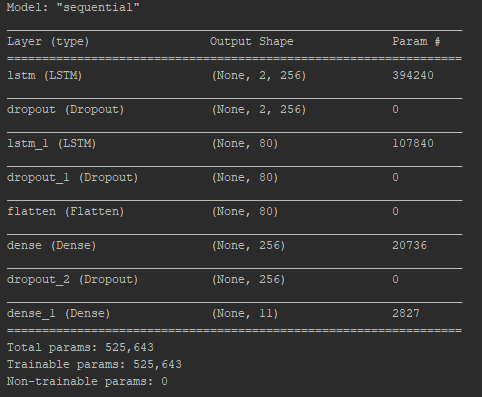
**Overview**

Our research determined that it was best to implement the LSTM type of RNN. Due to not having our own captured data, we resorted to using a publicly available dataset that fit our needs to train our model.

**Details**

The dataset ([RML2016.10a.tar.bz2](http://opendata.deepsig.io/datasets/2016.10/RML2016.10a.tar.bz2)) we used was gathered from the Deepsig website. The shape of the dataset is (220000 x 2 x 128), where 220,000 is the number of radio frequencies in the dataset and (2 x 128) is the input shape. The dataset consists of 11 modulations: 8 digital and 3 analog. (i.e BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK, and PAM4 for digital modulations, and WB-FM, AM-SSB, and AM-DSB for analog modulations). Our models take in those radio frequencies as an input and classify them into one of the 11 modulations.

We used the Tensorflow Keras API to implement our RNN model. It uses a network called LSTM, which stands for Long-Short-Term-Memory. It’s a type of RNN that deals with recognizing patterns in sequences of data. In our model, there are two LSTM networks followed by a dropout so that the neural network doesn’t overfit. Then, flatten and dense layers are used to get the final output layer to 11 neurons. Below is a summary representation of the model:



In addition, we used the CUDNN neural network library to help us run our model using NVIDIA GPU, which increased the speed and accuracy of the model. The hardware we used to run our model is on NVIDIA GeForce GTX 1060 graphics card. We trained our model with 154,000 samples and validated it on 66,000 samples.

# Issues

1. **COVID-19**: This was not an issue we expected to come across. We were hit with the pandemic about midway through the semester. The issue was that we couldn’t meet, meaning we were not able to collect data. We had to use data from various sites, but I believe our results would have been better if we were able to collect our own data.
2. **Dividing up the work:** Due to closure of the campus parts of the project had to be scrapped due to the hardware being in two different locations. That meant there was less work for the team, especially the data team since we couldn’t capture the data ourselves. This was overcome with combining the teams and putting more effort into the machine learning aspect of the project.
3. **Undertraining the RNN model:** Our group ran into some issues while running the RNN model. The biggest one was the low accuracy after training it. This took some research but was overcome by adding more iterations since we were undertraining it. The outcome was much better and more consistent to what we were looking for in terms of accuracy.

# Conclusions and Recommendations

**Conclusion**

In conclusion, we found that despite not being able to capture our own data, we had successful results while using publicly available data. The LSTM RNN worked well for our purposes in the initial phases of the implementation of this project, but there is still more work to be done to achieve higher accuracy on our model.

**Future Work**

We have two main tasks that could be continued from our work. The first being collecting better data, we used data provided to us through research. This data was not precisely what we needed to train the neural network though. So we would need to collect our own data in order for us to train the network exactly how we want it. The second task would be to hone our neural network to the newly acquired data. With our own data, we could train the neural network. This will improve our accuracy.

# 

# 

# References

A. Popleteev, V. Osmani and O. Mayora, "Investigation of indoor localization with ambient FM radio stations," *2012 IEEE International Conference on Pervasive Computing and Communications*, Lugano, 2012, pp. 171-179.Edelen, A.L., and et al. Initial Experimental Results of a Machine Learning-Based Temperature Control System for an RF Gun. United States: N. p., 2015. Web**.**

H. Stratigopoulos and Y. Makris, "Error Moderation in Low-Cost Machine-Learning-Based Analog/RF Testing," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 2, pp. 339-351, Feb. 2008.

K. Youssef, L. Bouchard, K. Haigh, J. Silovsky, B. Thapa and C. V. Valk, "Machine Learning Approach to RF Transmitter Identification," in *IEEE Journal of Radio Frequency Identification*, vol. 2, no. 4, pp. 197-205, Dec. 2018.

Mikaeil, Ahmed Mohammed, et al. “Machine Learning to Data Fusion Approach for Cooperative Spectrum Sensing.” *2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, 2014, doi:10.1109/cyberc.2014.80.

R. D. Miller, S. Kokalj-Filipovic, G. Vanhoy and J. Morman, "Policy Based Synthesis: Data Generation and Augmentation Methods For RF Machine Learning," *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Ottawa, ON, Canada, 2019, pp. 1-5.

Roy, Debashri, et al. “Machine Learning in Adversarial RF Environments.” *IEEE Communications Magazine*, vol. 57, no. 5, May 2019, pp. 82–87. *EBSCOhost*, doi:10.1109/MCOM.2019.1900031.

Thilina, K. M., Choi, K. W., Saquib, N., & Hossain, E. (2013). Machine Learning Techniques for Cooperative Spectrum Sensing in Cognitive Radio Networks. *IEEE Journal on Selected Areas in Communications*, *31*(11), 2209–2221. doi: 10.1109/jsac.2013.131120

# 

# Team

**Nick Taormino - Data collection team (pre-COVID19), Documentation**

I am a senior Undergraduate at Towson University with a major in Software Engineering and a minor in mathematics. Experience in web development, object oriented programming, and database management using queries. Languages include: Java, C++, HTML, Javascript, SQL, Kotlin, Assembly, Visual Basic, F#, Scheme, and Python. I will be a part of the data team, our main task in the beginning is to work with the hardware that was acquisitioned to collect data accurately.

**Nahom Getaneh - Development of RNN model**

I am a senior undergraduate at Towson University, majoring in Computer Science, with a Software Engineering Track. I have experience working with different programming languages such as Java, Python and Javascript. I also have experience with mobile app development using Android Studio and React Native. Furthermore, I have taken courses such as Data Structures and Algorithms, Object Oriented Programming, and Database Management Systems, and also worked in different projects that relate to MySQL, machine learning, and web. For this project, I am going to be working under the machine learning team, developing a RFML model.

**Luis Ramos - Data Collection and documentation**

A senior Computer Science major with a Software Engineering track. Experienced with programming languages such as Java, and C++, and topics in the subject, including object-oriented programming. I am knowledgeable on web development tools such as JavaScript, HTML, CSS, PHP, and SQL. Completed relevant courses such as Software Engineering on which experience with Agile Software Development Process was gained. For this project, I will be working with the data team on the initial phase.

**Brad Wilfong - Machine learning team and data collection helper.**

I am a senior at Towson University majoring in Computer Science with a track in software engineering. I have proficient knowledge in Java, HTML, CSS and Javascript with some experience with Kotlin, PHP, JQuery, Python and SQL/SQLite. I have completed projects that involved creating mobile applications using api’s, web app using sockets for chatting, web app using firebase to hold information and a machine learning project using Python. I have also worked in a team environment on most projects using applications like Slack and Github to collaborate.

**Nathan Cox - Development of RNN model and model training**

A senior Computer Science major at Towson University with 5 years of experience in Java and partial training in other languages such as Python, C++, C, and Javascript. Developed 2 mobile applications for Android and several skills for Amazon Alexa. Familiar with the agile development process and working in development teams.

**James Hooper - Data Collection Team and Documentation**

I’m an Undergraduate Software Engineering student expecting to graduate in May 2020. I have experience in the research process as I contributed to many research projects over the semester here at towson. I also have done a project that was associated with machine learning that was able to predict if data was functional or nonfunctional last semester for a group final project. I am excited to be contributing to this project and look forward to providing an innovative solution.