Computer Networks 199 (2021) 108441

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Data Article

Dataset for modulation classification and signal type classification for multi-task and single task learning

Anu Jagannath *, Jithin Jagannath *

Marconi-Rosenblatt AI/ML Innovation Laboratory, ANDRO Computational Solutions, LLC, Rome, NY, USA

ARTICLE INFO

Keywords:
Multi-task learning
Machine learning
Deep learning
Modulation classification
signal classification
software defined radio
Radio frequency dataset

ABSTRACT

Wireless signal characterization is a growing area of research and an essential tool to enable spectrum monitoring, tactical signal recognition, spectrum management, signal authentication for secure communication, and so on. Recent years have witnessed several deep neural network models to perform single task signal characterization such as radio fingerprinting for emitter identification, automatic modulation classification, spectrum sharing, etc. However, with the emergence of 5G and the prospects of beyond 5G communication, there has been an increased deployment of edge devices that requires lightweight neural network models to perform signal characterization. To this end, a multi-task learning model that can perform multiple signal characterization tasks with a single neural network model has been proposed. However, due to the novel nature of multi-task learning as applied to signal characterization, there is a lack of a corresponding dataset with multiple labels for each waveform. In this paper, we openly share a synthetic wireless waveforms dataset suited for modulation recognition and wireless signal (protocol) classification tasks separately as well as jointly. The waveforms comprise radar and communication waveforms generated with GNU Radio to represent a heterogeneous wireless environment.

Specifications Table

Value of the Data

- Why are these data useful?

Applied machine learning has been a key enabler for advancing various aspects of wireless communication recently [1, 2] and open datasets are critical to any supervised learning problem. Wireless signal characterization plays a primary role in elucidating the ongoing transmissions in the spectrum. Typically, such spectrum captures comprise heterogeneous wireless transmissions such as typical communication waveforms, spurious transmissions, etc. Most open wireless datasets are confined to typical communication waveforms. It is important to consider a wide range of waveforms representing heterogeneous communication environment. For example, spectrum sharing in the newly relieved spectrum bands such as 3.5 GHz and 6 GHz require the harmonious coexistence of unlicensed users with licensed incumbents. The incumbents comprise satellite and radar transmissions. Hence, it is important to consider radar waveforms in addition to the typical communication waveforms. Additionally, the dataset includes waveforms subject to realistic propagation and radio hardware uncertainties under varying SNR conditions. This is the only known wireless dataset suited for multi-task learning [3] which can be used for separate single signal characterization tasks as well.

- Who can benefit from these data?

Wireless researchers who are involved with applied AI/ML and spectrum managers who perform spectrum monitoring and management tasks.

- How can these data be used for further insights and development of experiments?

Spectrum managers and wireless researchers can evaluate their spectrum monitoring tool's capability to detect wide range of radar and communication waveforms with our dataset. Our dataset can be used to train, validate, and evaluate single task as well as multi-task signal characterization neural network models.

* Corresponding authors.
E-mail addresses: ajagannath@androcs.com (A. Jagannath), jjagannath@androcs.com (J. Jagannath).

https://doi.org/10.1016/j.comnet.2021.108441
Received 1 February 2021; Received in revised form 4 July 2021; Accepted 26 August 2021
Available online 2 September 2021

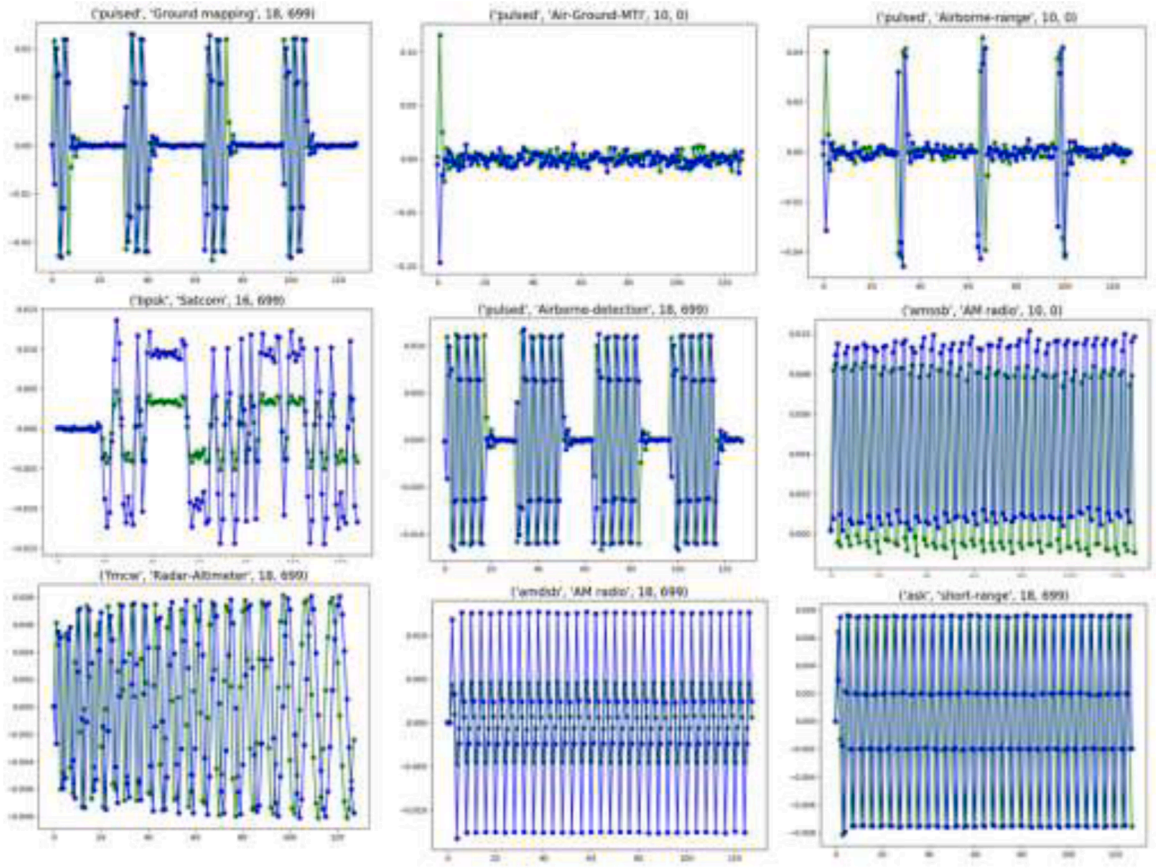

Fig. 1. Distinct waveforms in the dataset. The top label of each plot shows its corresponding key in the dataset.

Table 1
Dynamic parameter settings

Parameter	Value
Carrier frequency offset std. dev/sample	0.05 Hz
Maximum carrier frequency offset	250 Hz
Sampling rate	10 MS/s
Sample rate offset std. dev/sample	0.05 Hz
Maximum sample rate offset	60 Hz
Number of sinusoids in frequency selective fading	5
Maximum doppler frequency	2 Hz
Rician K-factor	3
Fractional sample delays comprising power delay profile (PDP)	[0.2,0.3,0.1]
Number of multipath taps	5
List of magnitudes corresponding to each delay in PDP	[1,0.5,0.5]
Range of SNRs in steps of 2 dB	-20 dB to 18 dB

- What is the additional value of these data?

Our dataset comprises waveforms subject to radio hardware defects and channel effects that are typical of realistic wireless environment. Most work in this field resort to identifying the modulation format of the waveforms [4, 5, 6, 7]. However, waveforms from multiple wireless standards can have the same modulation format and a single wireless standard could have multiple modulations. Hence, in addition to identifying modulation scheme our dataset allows for recognizing the wireless signal class as well. This is enabled with the multiple labels per example in the dataset. For more details, the reader may refer [3].

Data

The dataset is contained in an HDF5 container with key-value associations per entry. Each waveform is generated under dynamic effects over SNRs ranging from -20 dB to 18 dB. For each SNR, at least 659

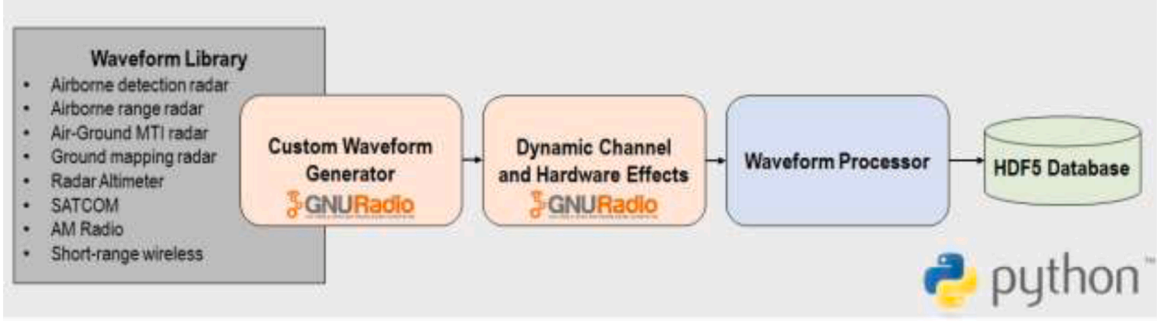

Fig. 2. Dataset collection framework

examples not exceeding 700 are created to provide a randomized sample set. Each example corresponding to a waveform can be accessed with its key ($id = \{modulation\ format,\ signal\ class,\ SNR,\ sample\ number$). The dataset collection process is shown inFig. 2.

Experimental design, materials, and methods

The waveforms are generated in software with Python scripts utilizing the GNU Radio libraries on an Ubuntu 18.04 VM running on an Intel Core i5-3230M CPU. The dataset collection process utilizes the dynamic channel block [8] in GNU Radio - gnuradio.channels.dynamic_channel_model.

Declaration of Competing Interest

Authors do not claim any conflict of interest for this submission.

Acknowledgments

(a) Contractor acknowledges Government's support in the publication of this paper. This material is based upon work supported by the US Army Contract No. W9113M-20-C-0067. (b) Any opinions, findings and conclusions or recommendation expressed in this material are those of the author(s) and do not necessarily reflect the views of the US Army.

References

[1] J. Jagannath, N. Polosky, A. Jagannath, F. Restuccia, T. Melodia, Machine Learning for Wireless Communications in the Internet of Things: A Comprehensive Survey, Ad Hoc Networks (Elsevier) 93 (June 2019), 101913.
[2] N.Polosky Jagannath, A. Jagannath, F. Restuccia, T. Melodia, Neural Networks for Signal Intelligence: Theory and Practice. Machine Learning for Future Wireless Communications (F. Luo, ed.), Wiley - IEEE Series, John Wiley & Sons, Limited, 2020.
[3] A. Jagannath, J. Jagannath, Multi-task Learning Approach for Automatic Modulation and Wireless Signal Classification, in: Proc. of IEEE International Conference on Communications (ICC), Montreal, Canada, June 2021.
[4] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M.M. Sebdani, Y. Yao, Modulation classification based on signal constellation diagrams and deep learning, IEEE Trans. Neural Networks and Learning Syst. 30 (3) (2019) 718–727.
[5] T.J. O'Shea, T. Roy, T.C. Clancy, Over-the-air deep learning based radio signal classification, IEEE Journal of Selected Topics in Signal Processing 12 (1) (2018) 168–179.
[6] J. Jagannath, N. Polosky, D.O. Connor, L. Theagarajan, B. Sheaffer, S. Foulke, P. Varshney, Artificial Neural Network based Automatic Modulation Classifier for Software Defined Radios, in: Proc. of IEEE Intl, Conf. on Communications (ICC), (Kansas City, USA), May 2018.
[7] R. Zhou, F. Liu, C.W. Gravelle, Deep learning for modulation recognition: A survey with a demonstration, IEEE Access 8 (2020) 67366–67376.
[8] GNU Radio dynamic channel models. https://www.gnuradio.org/doc/sphinx-3.7.2/channels/channels.html (accessed 31 January 2021).

Anu Jagannath received her MS degree (2013) from State University of New York at Buffalo in Electrical Engineering. She is currently a Senior Scientist at ANDRO Computational Solutions, LLC and serves as the Associate Director of Marconi-Rosenblatt AI/ML Innovation Lab at ANDRO. Her research focuses on MIMO communications, Deep Machine Learning, Reinforcement Learning, Adaptive signal processing, Software Defined Radios, spread spectrum systems, LPI/LPD communications, spectrum sensing, adaptive Physical layer, and cross layer techniques, medium access control and routing protocols, underwater wireless sensor networks, signal intelligence and so on. She serves as author and coauthor for book chapters and multiple research publications in journals and conference proceedings. She has rendered her reviewing service for conferences such as IEEE Annual Consumer Communications \& Networking Conference (CCNC) and IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). She is the co-Principal Investigator (co-PI) and Technical Lead in multiple Rapid Innovation Fund (RIF) and SBIR/STTR efforts involving developing embedded MIMO solutions, deep and reinforcement learning for wireless communications, signal intelligence, and mesh networking.

Jithin Jagannath received the B.Tech. degree in Electronics & Communication Engineering from University of Kerala, India; M.S. degree in Electrical Engineering from University at Buffalo, The State University of New York; and received his Ph.D. degree in Electrical Engineering from Northeastern University. Dr. Jagannath currently serves as the Director of the Marconi-Rosenblatt AI/ML Innovation Lab at ANDRO Computational Solutions. He is also the Adjunct Assistant Professor in the Department of Electrical Engineering at the University at Buffalo, State University of New York. Dr. Jagannath heads several of the ANDRO's research and development projects in the field of signal processing, cognitive radio, cross-layer ad-hoc networks, Internet-of-Things, and machine learning. He has been the lead author and Principal Investigator (PI) of several multi-million dollar research projects. This includes a Rapid Innovation Fund (RIF) and several Small Business Innovation Research (SBIR)s with objectives of designing and developing novel solutions in the domain of intelligent wireless networking, machine learning, MIMO wireless communication, and signal detection and classification among others. He is currently leading several teams developing commercial products such as SPEARLink™, DEEPLink™, DEEPSpec™, and ARROWLink™. He is an IEEE Senior member and serves as IEEE Industry DSP Technology Standing Committee member. Dr. Jagannath's recent research has led to several peer-reviewed journal and conference publications. He is the inventor of 7 U.S. Patents (granted, pending). He also renders his service as Editor, TPC Member and Reviewer to several leading journals and conferences.