Analysis of Langmuir probe CubeSat data with a neural network[a)]

Matthew Lazo[b)]
Department of Physics, West Virginia University, Morgantown, WV, 26506, USA

(Dated: 5 November 2020)

Abstract: Plasmas across a range of applications are often analyzed with Langmuir probes, which sweep a bias voltage across the probe tip at several kHz and produce a large volume of data to be processed for each current-voltage characteristic. Traditionally, each probe trace is manually truncated and fitted to derive relevant plasma parameters such as temperature and density. Standard Langmuir probe theory is well-understood for Maxwellian plasmas, however, and probe characteristics can be simulated as a piecewise function with noise. Therefore, an opportunity arises to build an automated, neural-network based workflow for extracting plasma parameters. The NN is trained on simulated space plasma data, then trained further on real, human-labelled space plasma data. The network's ability to diagnose previously unseen space plasma data is then tested and compared with human analysis.

I. BRIEF BACKGROUND

The theoretical background for Langmuir probes and neural networks are drawn from several sources[1,2,3,4,5,6], and will be discussed in detail in the final report. Only a brief discussion is given here to establish the ideas and plans of the project in this short progress report.

The neural network implemented here was originally built by the author as part of a Department of Energy Student Undergraduate Laboratory Internship project with the Princeton Plasma Physics Laboratory over the summer of 2020, and full technical specifications of the neural network structure are contained within the report by Lazo et. al[7].

That previous project's neural network shall be re-purposed to function with a different "flavor" of data. The original network[7] was built to operate on fusion reactor plasmas within the National Spherical Tokamak Experiment, where temperatures often reach $25eV$ or more, with particle densities on the order of 10^{20} particles per cubic meter. In this regime, the *plasma sheath* does not change once it appears, and electron saturation is seldom attained. This project will attempt to analyze space plasma data, particularly Langmuir probe data taken from the Simulation-to-Flight 1 CubeSat orbiting at an altitude of approximately $500km$ above the Earth, within the magnetosphere where plasmas blanket the atmosphere. Typical densities of singly-ionized, monatomic oxygen at the CubeSat's altitude are in the realm of 10^{16} particles per cubic meter, with electron temperatures ranging between $0.1eV$ and $5eV$[8].

A. Langmuir Probes

Invented by Irving Langmuir, Langmuir probes are the simplest way to measure a hot, ionized gas, or plasma[2]. The simplest probes consist of a thin metal cylinder or length of wire inserted into the plasma, and given a bias voltage relative to the electrostatic potential

---

[a)]Internal report on final project progress for the undergraduate computational physics course at West Virginia University

[b)]Electronic mail: mjl0018@mix.wvu.edu

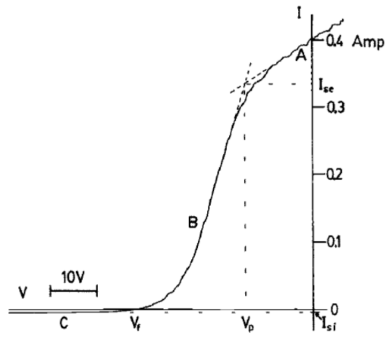

FIG. 1. Example Langmuir probe trace, taken from Hutchinson[1]. The three regions of ion saturation, transition, and electron saturation can be seen.

of the plasma. The bias voltage is swept in time, often several hundred times per second, from a negative to a positive potential relative to the plasma[1]. Being made up of charge-carrying electrons and ions, particles in the plasma will be attracted to or repelled from the probe tip, and corresponding ion and electron currents are collected as the probe sweeps from negative to positive voltage. The resulting current-voltage characteristic, or I-V trace, is then analyzed to determine plasma parameters such as electron temperature and species density or energy distributions. The total current collected by the probe, according to a classical, kinetic treatment of the plasma, is given by Hutchinson as

$$I = neA_p \left(\frac{T_e}{m_i}\right)^{1/2} \left[\frac{1}{2} exp\left(\frac{eV_b}{T_e}\right) - \frac{A_s}{A_p} exp\left(-\frac{1}{2}\right)\right], \tag{1}$$

where n is particle density – taking electron and ion densities to be equal for a quasi-neutral plasma – e is the fundamental charge, A_p is the surface area of the probe, A_s is the surface area of the plasma sheath around the probe, T_e is electron temperature in electron volts, V_b is the bias voltage in volts, and m_i is the ion mass in kilograms. The energy distributions of ion and electron species are taken to be Maxwellian. The ratio A_s/A_p depends on the plasma setting, and is likely to vary in time for the space plasma considered here, although this must still be determined. Additionally, we note that we shall refer to Ruzic[2] in deriving the appropriate theoretical model, based on Eq. (1), to accurately describe the data we will use, and this model will be explained in-depth in the final report.

B. Neural Networks

Machine learning, particularly deep learning, is built on neural networks. Langmuir probe data evolves in time, and so the data we will work with shall consist of time series. We believe that plasma time series data can be appropriately modeled by employing a Long Short-Term Memory recurrent neural network[5]. In brief terms, a neural network is a large function consisting of nodes which hold numerical values. Data passed into the network passes through several layers of nodes, and is associated with a unique statistical weight as each value passes from one node in one layer to each node of the subsequent layer. By the end of the network, the node values have been adjusted by each weight in such a way that a decision can be made about the input data. The output is either a classification (say, an image of a dog) or prediction of an optimal regression parameter (i.e., an optimal slope for a linear fit). These functions often consist of multiple layers and hundreds of thousands of nodes. The process of "learning" undergone by a network is done by rapidly passing new data sets through the network and allowing it to make predictions, while also supplying the correct answers so that the network can judge how "good" its estimate is, and adjust the internal statistical weights to make a better guess next time.

A more in-depth description of this topic will be provided in the final report, and is also treated by Lazo et. al[7]. The project utilizes a machine learning library, and so the network is not built from the ground up.

II. PLANNED PROGRAM STRUCTURE

The program is implemented in the Python programming language and utilizes Tensorflow 2.3.0[9], an open-source machine learning library developed by Google. Deviations from the original network design[7] will be noted in the final report, although at the time of writing the same structure is implemented with a more widely recognizable assessment of training loss: four bidirectional LSTM layers are followed by three standard Dense layers which use, respectively, a sigmoid activation function and two hyperbolic tangent activation functions[8]. The network is compiled with the Adam[6] optimization algorithm and uses the coefficient of determination[10] method to describe the error between network training predictions and provided labels.

In order to fully diagnose plasma parameters, several operational stages are required. A main script imports and calls functions from several other scripts to execute each stage. The necessary scripts each handle a specific task: data import and formatting; calculation of plasma parameters from real data; construction of synthetic probe data based on the aforementioned theoretical model; and construction and compilation of the neural network, followed by execution of two training stages. At the time of writing, the code necessary to build the neural network has been written, and the code to generate synthetic data must be adapted from existing code built with fusion reactor plasma parameters in mind[7].

The actual implementation of this structure will take place consistently over several weeks, with a possible implementation of the Google Drive API to automate the extraction and pre-processing of data. Data is downlinked from the CubeSat to NASA Internal Validation and Verification, where it is processed and uploaded to a Google Drive folder. The scientific data extracted from this is combined from NORAD satellite tracking data to build a profile of the plasma temperature and density around the Earth with the purpose of observing day-night cycles of plasma behavior. This project will seek to cut out many hours of manual human processing in determining these temperatures and densities. The process of plotting and determining day-night plasma cycles shall remain out of the scope of this project.

III. REFERENCES

[1] I. H. Hutchinson, *Principles of Plasma Diagnostics* (Cambridge University Press, 2002) pp. 55–72.
[2] D. N. Ruzic, *Electric Probes for Low Temperature Plasmas*, M-13 (The American Vacuum Society Education Committee, 1994).
[3] F. F. Chen, *Introduction to Plasma Physics and Controlled Fusion*, 3rd ed. (Springer International Publishing, 2016).
[4] D. S. Thompson, "Three-dimensional multispecies distribution functions in a plasma boundary with an oblique magnetic field," dissertation (2018).
[5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation **9**, 1735–1780 (1997), english translation.
[6] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," (2015), iCLR 2015 Conference Paper.
[7] M. J. Lazo, X. Zhang, and F. Poli, "Langmuir probe data interpretation with a neural network," (2020), dept. of Energy internal report.
[8] Y. Kamide and A. C.-L. Chian, *Handbook of the Solar-Terrestrial Environment* (Springer-Verlag Berlin Heidelberg, 2007) pp. 222–242.
[9] https://www.tensorflow.org/, website and documentation (2020).
[10] https://en.wikipedia.org/wiki/Coefficient_of_determination, "Coefficient of determination," Wikipedia article (2020).