Comparing Optimization Techniques on GEDI LiDAR Waveforms

Andrew C. Baker, Arjun Raj, Sara I. Ansari
Under the mentorship of:
Brandon J. Woodard, Department of Computer Science, Brown University
Shinpei S. Nakamura, Department of Statistics and Data Science, Yale University

Abstract

In our study we aim to compare the performance of Stochastic Gradient Descent (SGD) (Sun, et al. 2019) and the Adam optimization (Sun, et al. 2019) algorithms in an effort to determine which better predicts canopy height. We will utilize the data made available through the Global Ecosystem Dynamics Investigation's (GEDI) LiDAR technology to generate waveforms (Global Ecosystem Dynamics Investigation (GEDI) n.d.). We will then train models on predicting canopy height using the Level 1B data which consists of raw GEDI waveforms displayed as amplitude as a function of time. We will then apply the aforementioned optimization methods to these models and compare their performances based on a multitude of factors. This research will assist ecologists, biologists, and earth scientists in creating more accurate biomass models, in better understanding important carbon processes, as well as address issues such as climate change, ecological health, and carbon emissions.

Introduction

The study we propose will compare the performance of two optimization function's ability to accurately model GEDI data. GEDI or NASA's Global Ecosystem Dynamics Investigation was scheduled for a two year deployment on the International Space Station (ISS) in 2018, but the deployment was extended to retrieve more data. GEDI produces high resolution observations of the three dimensional structure of the Earth's topography using its full form LiDAR instrument. LiDar is a technology that uses laser pulses to measure structures, spaces, and surfaces. GEDI data is of great value for studies about forest and water resource management, carbon life cycle science, and weather prediction.

The two optimization functions we propose using are fairly similar. Stochastic Gradient Descent is an effective optimization technique that limits the computational burden of normal optimization functions which increases it's time efficiency. On the other hand, the Adam optimization function is a newer method which is intended to be a replacement for Stochastic Gradient Descent. Adam excels in handling data with sparse densities and a lot of noise. We intend to determine if Adam is the more viable optimization function for working with LiDAR waveform data.

Methods

We will create two separate models that predict canopy height utilizing GEDI Level 1B data. One model will use the stochastic gradient descent optimization function, and another will use the Adam optimization function. We will then apply both models to a testing set of data and record the following measurements:

1. Time the model takes to generate
   a. Measured in seconds (s).
   b. We will measure the time taken to complete the process of optimizing by using python's time module or the data training library's built-in functions.
2. Memory Usage
   a. Measured in GigaBytes (GB)
   b. We will measure the memory using the os and psutil libraries functions.
3. Central Processing Unit (CPU) Usage
   a. Measured in CPU usage percentage (Percent of total CPU power used)
   b. We will measure the CPU usage using the os and psutil libraries functions
4. Accuracy
   a. Measured in a percentage
   b. The accuracy will be measured using the data training library's built-in functions.

We will test each variable 100 times and take the average of our tests. We will then compare the measurements in order to determine which is a more viable optimization function. Additionally, we will be evaluating whether or not the results we obtain are statistically significant through the use of p-values.

Timeline

Week 1 - Introduction to Computer Science research & LiDAR technology

Readings:

1. Computer Science (What is CS?) [Toolsmith-CACM.pdf (unc.edu)]
2. The GEDI Simulator: A Large-Footprint Waveform LiDAR Simulator for Calibration and Validation of Spaceborne Missions - Hancock - 2019 - Earth and Space Science - Wiley Online Library (Sun, et al. 2019)[https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018EA000506]

Week 2 - Learning how to read a scientific paper and becoming familiar with data modeling in Python

Readings:

1. How to read a paper [http://blizzard.cs.uwaterloo.ca/keshav/home/Papers/data/07/paper-reading.pdf]
2. Set-up Python Environment for GEDI data: https://youtu.be/UlrCC1Xp-wk

3. Referring and learning from a virtual textbook for learning Python and applying it to data science [Python Data Science Handbook | Python Data Science Handbook (jakevdp.github.io)]

Week 3 - Model a waveform in Python and transitioning into machine learning/AI topics

Readings:

1. Paper on machine learning [https://stacks.stanford.edu/file/druid:jt687kv7146/jt687kv7146.pdf]
2. Optimization functions [https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html#scipy.optimize.least_squares https://docs.scipy.org/doc/scipy/reference/tutorial/interpolate.html]

Week 4 - First machine learning project

1. Following and Implementing all exercises in the Tech with Tim machine learning video series [https://youtu.be/ujTCoH21GlA]

Week 5 - Proposal Writing

Readings:

1. A proposal made by Brandon Woodard for Brown University's Computer Science department.

Writing:

1. Outline of Proposal: Sections and content

Week 6 - Progress report

Review, revision, and feedback of proposal. Address changes for next iteration

Week 7 - Create Model

1. Watch through two B Spline Interpolation Videos:
   [https://www.youtube.com/watch?v=dxvmafuP9Wk]
   [https://youtu.be/83Nsel2uv6k]
2. Create Models

   Readings:

   1. Read through the python B spline libraries and familiarize yourself with them. [scipy.interpolate.BSpline — SciPy v1.6.1 Reference Guide]

Week 8 - Run Tests

1. Time Tests
2. CPU Usage Tests
3. Memory Usage Tests
4. Accuracy Tests

Week 9 - Begin Analyzing data/Begin Writing Paper

1. Watch through these two S
2. Write Beginning of the Paper

Week 10 - Complete First Draft of Paper

Week 11 - Review and Revise Final Paper

Week 12 - Wrapping up final paper

1. Final Paper submitted by the end of the week

References

n.d. *Global Ecosystem Dynamics Investigation (GEDI).* https://gedi.umd.edu/.

Sun, Shiliang, Zehui Cao, Han Zhu, and Jing Zhao. 2019. "A Survey of Optimization Methods from a Machine Learning Perspective." *Arxiv* 30. https://arxiv.org/pdf/1906.06821.pdf.