A Novel Neural Network and Application to Infer Soil Organic Carbon for Carbon Sequestration in Agricultural Soil

Background

Introduction

As of 2015, the global mean surface temperature has increased 0.87 degrees Celsius relative to 1850-1900 levels due to climate change [1]. Soil is one of the largest carbon sinks, storing 2,500 billion tons of carbon, more than the atmosphere and plant life combined[5]. Global agricultural soil has lost 50-70% of its original soil carbon through agricultural practices[5]. Carbon sequestration through soil can mitigate global climate change by increasing soil organic carbon (SOC) in agricultural soil[2]. Some agricultural practices have been associated with increasing SOC, including minimum tillage and crop covering. The effectiveness of these practices, however, is usually specific to individual farms[2]. Farm-specific evaluation of these practices is difficult as SOC cannot be measured directly [6] and frequently can only be measured in a lab with high-resolution soil data[3]. Existing analytic methods to calculate SOC are inconsistent between labs due to high analytic complexity[3]. Without real-time SOC data, it may be difficult for farmers and researchers to identify how certain practices affect carbon sequestration in agricultural soil. Similarly, it could prevent farmers from changing their farming practices in response to live SOC data.

Solution and Application

A new machine learning method of inferring SOC in real-time using existing technology and data could allow farms to adapt their soil management practices to increase carbon sequestration. A neural network could use soil databases to train a machine learning algorithm to infer SOC. Farmers and researchers can use inferred SOC data to identify what affects SOC on an individual farm and modify farm practices in real-time to sequester carbon. Software to infer SOC and modify farm practices can also improve soil quality and productivity, as carbon improves soil structure, water-retention, and fertility by increasing the content of soil organic matter[6]. A similar solution is to use new soil spectroscopy technology. This technique is limited as it requires non-existent or limited soil spectroscopy data and available commercial products are more limited than traditional soil sensors[3].

Design Criteria

For maximum adoptability by farmers, the neural network should use inputs that are easily measured by sensors. If farmers must install new physical sensors with the software it may decrease incentive to track carbon sequestration. The software should also be easy to use and eventually portable across common operating systems. The initial neural network should be accurate to within half a standard deviation so it can identify long-term trends in data.

Training Data

For a machine learning algorithm to be accurate, it is imperative that the data is accurate and come from a reliable source. The data used to train the neural network is from the USDA's National Cooperative Soil Survey (NCSS), a partnership between government agencies and private institutions to record soil data using common standards across the country[4]. This data was chosen due to its wide variety of measurements which could be incorporated into the neural network and because it focuses on measuring with common standards.

Design and Development

Goals

1. A mean absolute error of within half a standard deviation of the data.
2. Inputs that are simple to measure with sensors and on the field.
3. Software contains features to analyze SOC against other variables and events.
4. Software can track trend of SOC over time for a set of sensors.
5. Software can find SOC in real-time.

Design Timeline

Identify Dataset

The NCSS databaase was chosen to train and test the neural network for its reputability and wide variety of possible input variables.

Reserach Alternative Solutions

Soil spectroscopy is an alternative to machine learning for live soil monitoring. Spectroscopy, however, has a greater barrier to entry for most farmers because it frequently requires new equipement. There is also a lack of soil spectral data to make accurate preditions in many cases.

Segment Dataset

Around 20 variables which could influence SOC were chosen from the database. Approximately 200 samples from Fresno County were selected.

Train and Test Model

A basic neural network was trained and tested using the selected data. The mean absolute error was inedequate and failed to reach the goal.

Design Iteration

Because 200 samples may not have been enough data to adequately train the model, data from the entire state of California was selected. Variables with more available data were also chosen.

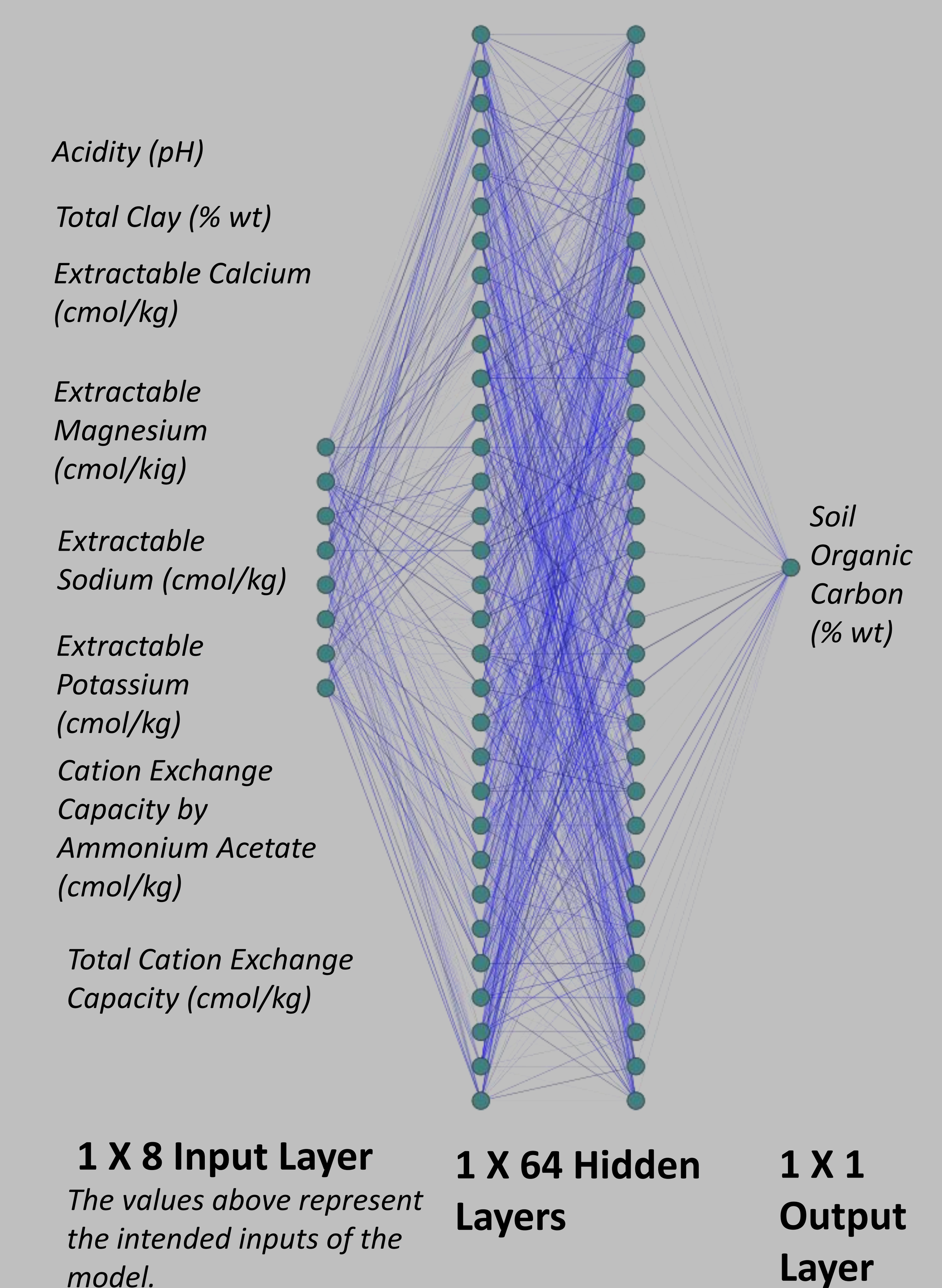
Model Optimization

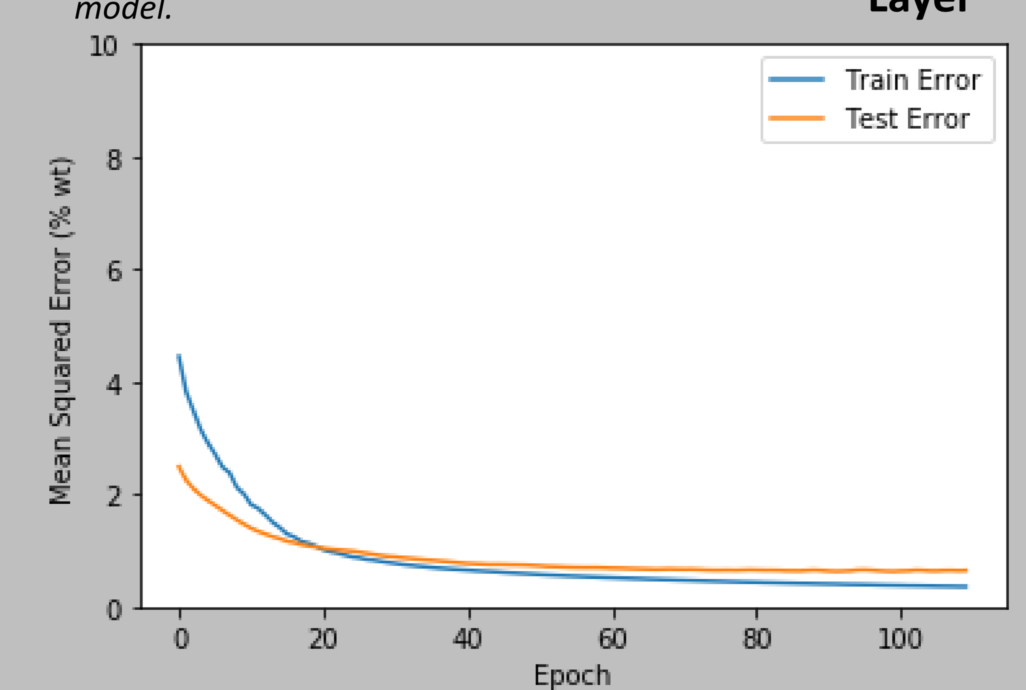
The dimensions of the neural network and weight optimizer were modified to improve the accuracy of the model. The model uses the ReLU activation function and Adamax[7] optimizer. An early-stop feature was added to the training stage to prevent the training error and test error from diverging. This is known as "overfitting" and can harm the accuracy of a neural network.

Desktop App Design

The technologies React and Electron were chosen to create a desktop application to use the neural network. These frameworks allow the application to use the Chromium rendering engine to be fully modular across Linux, Windows, Mac OS, and even as a website.

Neural Network Architecture





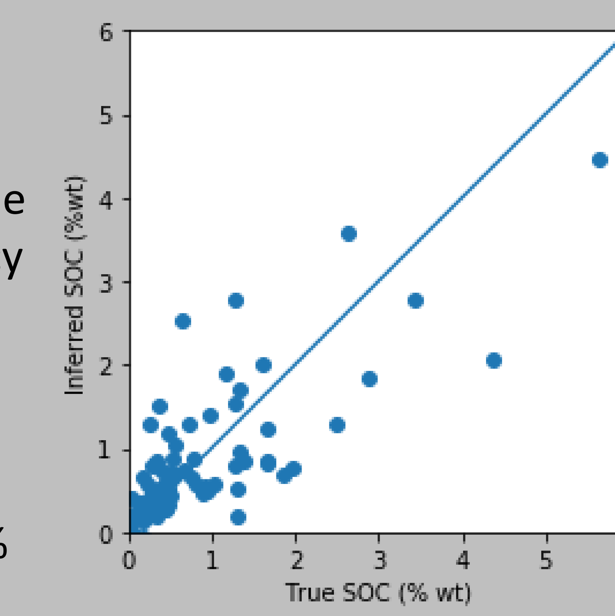
Error of the neural network as it was trained. Train error is the loss that was used to optimize the weights. Test error did not affect training and represents the true error of the model. An "epoch" is one iteration of changing the weights.

Desktop Software Design

The desktop application was developed using the Electron framework with the JavaScript programming language. The framework allows the application to be rendered with the Chromium engine as if it was a website. The application was built with the HTML, CSS, and JavaScript programming languages while the neural network itself was built with Python. It uses TensorFlow to run the trained model. The software was developed on the Linux operating system but was designed to eventually be fully modular on Linux, Windows, and Mac OS.

Evaluation

Accuracy



Final mean absolute error: 0.59 % weight of SOC.

20% of the data was used to test the accuracy of the model. The accuracy goal was approximately reached. The model is accurate within about 0.6 % weight of SOC (not 0.6 % error). The mean and standard deviation of the dataset was 1.45 % weight and 3.78 % weight, respectively. This places the accuracy to within one half of a standard deviation.

Limitations and Future Improvements

Although the accuracy goal was approximately reached, more thorough testing must be done across different geographic locations. The data used to both train and test the neural network was concentrated throughout California. Because few counties in California individually had enough data with the necessary variables, the performance of the neural network could not be tested in small areas. Similarly, the neural network was not tested with data from other states and geographic locations. The neural network was also not tested using new field data or sensor data.

An example of a soil pH meter from Spectrum Technologies , Inc.

The values used to find the SOC can be measured relatively easily with existing meters. However, the values for extractable magnesium assume little presence of $MgCO_3$. This is because the NCSS measured extractable magnesium using a method that does not count magnesium in $MgCO_3$. Although sensors for each of the variables exist, their accuracy has not been tested and it is unclear which sensors farmers typically have.

Future Goals and Plans

1. Test neural network with newly sampled soil and live sensor data.
2. Identify which sensors farmers already have. Build new neural networks to use data from those sensors.
3. Make the desktop application simpler to use and modular on Windows and Mac OS.
4. Integrate the application with ThingSpeak to read data directly from sensors.
5. Create a plug-in or feature to use the neural network with existing farm monitoring software.