Problem Statement: Coal and Open-pit surface mining impacts on American Lands Follow-On (COAL-FO)

Bryce Egley and Kenny Thompson
CS 461: Senior Capstone

October 17, 2017

Abstract

In 2016-2017, three computer science seniors from OSU developed software to detect the damage caused by coal mining. Their work was impressive, they developed a suite of algorithms designed to detect and track damage using satellite imagery. They even managed to win an award and get mentioned in the OSU alumni newspaper. However there were several promising opportunities to use their work that was uncompleted at the time of graduation. Our project, and our goal, is to take their work, improve upon it, and share it for the world to see. We will be doing it by taking their work and their algorithms, and converting them into tools that will run on the XSEDE cluster. We will run it on the XSEDE cluster, feeding it Terabytes of publicly available spectrometer readings collected from the NASA AVIRIS project. We will then take that data, and create a search engine to view the processed data in a readable form. People have created these tools before, and done these analysiss before, but it has not been created and made publicly viewable online. This will make future projects easier for researchers in this field, and will create new and exciting opportunities for scientific analysis.

Definition and Description

This project will create new as well as improve upon existing methods to classify, characterize and quantify impacts of mining and other destructive surface mining activities across the United States. The COAL project will now go on XSEDE, a single virtual system where individuals will be able to share computing resources. By putting the COAL project on XSEDE others will be able to use the program and data that comes from the COAL project. This will also enable us to further improve the algorithm's, execution runtime performance and geospatial output results. The original COAL project focused on gathering data from remote sensing devices, providing a suite of algorithms for classifying land cover, identifying mines and other graphical features, and correlating them with environmental data sets as described on the coal capstone github page.
This is the fourth capstone project and hundreds of people have forked the code from the previous project off github. Since, so many people are using the code produced in this project it is very important that we make images viewable on XSEDE so people who want to use the data from this project in the future will be able to easily gather the data they want and be able to apply it to their own projects.
Overall this project is very important since we need to track and monitor the environmental changes and effects coal has on surrounding lands since Coal is one of America's main sources of energy. We want to make sure we are extracting coal and using it safely for the sake of the environment. Satellite imaging using remote sensing devices allows us to gather much more data than a human eye is able to. Satellite imaging can provide a picture of an entire area of land and see things invisible to the human eye such as methane emissions.

Proposed Solution

So, our problem is, that to get this data which is essential to many research projects, researchers have to create tools and process the data, sometimes processing data thats already been done. The first half has already largely been solved, thanks to the 2016-2017 capstone team. The real issue is the second part. In order to solve the problem addressed. we wish to put the COAL-FO (COAL follow on) project on XSEDE, a high powered computing cluster, and analyze as much publicly available information as possible. XSEDE only allows for 2 terabytes so we will come up with a method to feed it data that will be looped through. The main priority for our capstone project is making a searchable port for the COAL project. We also plan on having meetings every two weeks where we will turn in and discuss past week deliverables and plan future deliverables. This is where we will create a schedule of how to complete the project. As already discussed we will also plan on improving the existing COAL architecture and systems. We also hope to improve the imagery processing algorithms currently in use. Then we will hope to create a new baseline suite to rank changes to areas of land over time to determine if the changes in

one particular area our significant or not.

Performance Metrics

The performance metrics for this capstone project will be to complete as many of the objectives as possible discussed in the abstract and the project description. Our main priority according to our client is to create a searchable port for the project. Once we complete this we will move on to the other objectives. We will also have bi-weekly performance metrics based on the meetings we have with our client and the deliverables that come from those meetings. If we can complete those objectives and the high priority objectives then this project will be a success.

This means we don't necessarily need to complete all the objectives for this project to be successful. As long as the high priority objectives like the searchable port and maybe putting the project on XSEDE are completed. The other objectives should only be given focus once we have completed work on the high priority objectives. These lower priority objectives would include improving existing algorithms and improving the accuracy of finding correlations between land changes in different images.

Overall this project looks challenging and it seems like a lot of work but if we focus on one objective at a time and use the resources available to us I believe we can accomplish several of the high priority objectives. We also hope that the work we do on this project will make it more accessible and usable for other individuals working on their own projects and research. We look forward to working with the NASA Jet Propulsion Lab as well as the other members of the COAL project from previous years.