INFX 573 – Data Science I
Data Science Seminar Response Paper
Charudatta Deshpande
12/7/2017

Speaker – Kelsey Jordahl from Planet Labs (https://www.planet.com/)

Subject – Mosaicking the earth every day

Venue – University of Washington eScience Institute

Q1 - What did the presenter do? Try to stress the data/technical part in the talk.

The presenter, Kelsey Jordahl from Planet Labs, explained how their company takes satellite images of earth every day, and how the image data is used to build homogeneous and seamless images of the earth. This process is referred to as 'mosaicking' since bits and pieces from various images are combined together to create one seamless image that is the best possible result from all the images that were taken by the satellite. The data generation process is carried out by two constellations of satellite. One constellation is at International Space Station (ISS) level orbit, while the other one is much higher, at Sun Synchronous orbit. While this sounds easy, or not related to data science, there are actually multiple data intensive processes running behind the scene that make it possible. The presenter showed an example of a 'tile' of earth 20 km long and 20 km wide and explained how one such 'tile' is created from multiple images. The images are basically converted into pixel level data, and various validation methods are first applied. The methods analyze the data, and rank them according to the criteria specified by the data analyst. The criteria usually include better weather, less cloud cover, less snow, and image resolution. Once the images are ranked, they are mosaicked together to create one single tile which appears seamless to a human eye. This is just the starting point. Most tiles require a color correction, so they reflect the colors as they appear to a human eye. This is done using reference images, which come from a different source called 'Land Sat'. The reference image is converted into a histogram that identifies various colors in the image, and their proportions. The source image, the tile, is also converted into a similar histogram. Then a Continuous Distribution Function (CDF) process runs and tries to match the source histogram to the reference histogram as much as possible. The resulting image retains all original photographed features, but is now color corrected to how a human would see it. There are multiple other data intensive processes that accommodate weather modeling and other normalization methods. Right now the company can create weekly mosaics of the entire globe. The company aims to achieve a daily mosaic of the entire globe every day using advanced data analysis process, higher computing power and by increasing the number of satellites in the orbit.

Q2 - What were the main findings/results/product?

The final product of this processing is the images of globe that appear seamless to human eye, and provide best insights from millions of satellite images. These images are an effective way of doing time analysis of the entire globe, and detect the changes that happen over time. The presenter showed some 'before' and 'after' images of some places on earth which clearly illustrated the changes. E.g., 2016 and 2017 snapshots of a glacier in Tibet indicated that the glacier practically collapsed in 2017, which could be effect of global warming. Similar snapshots of Orville dam in California illustrated the damage caused by the flooding, and how the water was released from the dam which prompted the evacuations in the area. For informational purpose, the speaker also showed images of UW campus in various seasons and identified the buildings.

Such images have proven to be highly effective tool for scientists and academics, who are conducting research on topics like global warming, deforestations, weather modeling, draught analysis etc. The company provides access to scientists and academics to their data for a fixed period of time. For projects that run longer, appropriate lengths of access times are also available. The company recently achieved weekly mosaicking, which will also prove to be very effective way of analyzing short term changes, and also will provide an accurate history to study longer term effects.

Q3 - Your personal comments. Was it interesting? Do you think it is relevant? Would you mind working on something similar? How well did you understand?

I personally enjoyed the seminar. It was interesting for me, because earlier in 2017 sheer coincidentally I was closely following the launch that deployed 86 of their new satellites in the orbit at same time. The launch was carried out by Indian Space Research Organization (ISRO) which set a world record of sending 104 satellites in same launch. Out of those 104, 86 belonged to Planet Labs. The presenter shared a video of the deployment process, which I had already watched as part of my interest in activities of ISRO.

It is highly relevant to the topics covered in class. They use Cumulative Distribution Function (CDF) to match the color histograms of source and reference images. Also they use regression models to find best possible images out of millions of images sent by the satellites. All these topics have been discussed in class, and it was fun to see how they are applied in the industry.

I would really like to work in geo-spatial data analysis field like this. I believe it helps the human race as a whole, and helps us better understand our own planet. I am also passionate about things like global warming, deforestation etc. that affect us all, and are mostly caused by human activities. Projects like this help us better understand these phenomenon, and can be used as starting points in neutralizing ill-effects of these phenomenon.

As to how well I understood, honestly, the presenter did not go into much technical details of how they do CDF and regressions. He only spoke at a higher level, and provided example images, but never really went into minute technical level details of what was actually being done. I understood the part that he presented, but I believe the actual mathematical

models will be highly complex, and will require a much higher level of expertise to completely comprehend the technical level details.

Note – Due to logistical constraints, I was unable to attend a Data Science seminar in person. With Prof. Toomet's pre-approval, I reviewed an online version of the seminar conducted at University of Washington eScience Institute. The online version is available at

https://www.youtube.com/watch?v=uM2yqQI66F8&list=PLA6PlfxWZPLRSOMVdrE9AZ_FPDzX5bh2d&index=3