2017 Spring: CS5542 Big Data Analytics and Apps

Technical Background & Project Development

Last Name: Natesan Arumugam

First Name: Bharath Kumar

UMKC email: bn4n5@mail.umkc.edu

- Machine Learning (Scale 0 – 10): 0
- Deep Learning (TensorFlow, Caffe, etc) (Scale 0 – 10): 0
- Spark MLlib (Scale 0 – 10): 5
- Hadoop (Scale 0 – 10): 5
- Cloudera (Scale 0 – 10): 0
- Android App (Scale 0 – 10): 8
- Java (Scale 0 – 10): 7
- Scala (Scale 0 – 10): 7
- Python (Scale 0 – 10): 3
- JavaScript (Scale 0 – 10): 8
- Projects: Have you ever worked on a software development project? If yes, describe them.
    1. Robust Resume Maker - A web application to generate user's resume by mashing up of information obtained from social networking sites associated with them.
    2. Share Up – Grocery Management Application - An android application developed using ionic framework to assist users in finding cheap grocery deals.
    3. Twitter Data Analysis using Apache Spark - Developed a system to store, analyze, and visualize a social network's (e.g. Twitter's) data.
- Project Partners for Your CS5542 Project (if any)
    1. Last Name: Ramesh
       First Name: Sibi Chakravarthy
    2. Last Name: Naravula Loganathan
       First Name: Barath
    3. Last Name:
       First Name:

Indicate if you have your own devices. If yes, describe about your device.

	Yes/No	System Spec (Processor, Memory, Hard Drive, Graphic Card, Operating System, etc)
Laptop	Yes	Intel Core i7-6500U 2.50 with Turbo Boost Technology up to 3.10 GHz 8GB DDR3L 1600 MHz, 256 GB SSD, Windows 10/Ubuntu
Smartphone	Yes	Qualcomm® Snapdragon™ 810 v2.1, 2.0 GHz octa-core 64-bit Adreno 430 GPU, 64 GB, 3 GB RAM, Android 7.1.1
Desktop		

CS5542 Big Data Analytics and Apps

Problem Set-1 (PS-1)

Jan. 17 (T), 2017

Name: Bharath Kumar Natesan Arumugam

Email: bn4n5@mail.umkc.edu

1. **Question: Odd Man Out**
   You're given an unsorted array of integers where every integer appears exactly twice, except for one integer which appears only once. Write an algorithm (in a language of your choice) that finds the integer that appears only once.

2. **System Design: Big Data Applications**
   A group of primatologists wants to study the details of the daily movement, activities, and interactions of a group of 6 chimpanzees living on "chimp island" - a natural, though somewhat open habitat about 50 meters in diameter, bounded on all sides by water, in the San-Diego zoo. Since they don't want to sit all day every day recording the second-by second positions and activities of the chimps, they have come to you, a computer vision expert, for automated assistance. They are interested in both compiling statistics about the movement and location of individuals, and in the frequency and locations of different interactions and activities (feeding, sleeping, grooming, fighting, etc.) They are willing to help in labeling relevant activities, even to the point of answering a few hundred quick questions per day of data (what's she doing here?), but they don't want to sit through 24 hours of video to do it. Ultimately they want an automated database that they can use to find out how many hours a day chimp Jane sleeps and where, histogram preferred eating locations, obtain statistics on who grooms whom, etc. Their equipment budget is an exceedingly generous $50K. P.S. If you want to "bar-code" the chimps, you have to do it in a way that will bother neither the chimps, nor the visitors to the zoo.
   - What kind of data do you want to extract?
   - How to collect such data?
   - How big data analytics can uncover the unexpected in your data?
   - How does the learning improve your system?
   - Draw the workflow of the proposed system and explain the process of the system.

1.

```cpp
#include <iostream>
using namespace std;
int main()
{
int arr[n];
int x=arr[0];
for(int i=1;i<n;i++)
 {
 x=x^arr[i];
 }
cout<<"the number that appears once: "<<x;
return 0;
}
```

2.

What kind of data do you want to extract?

- Different images of Chimps and their activities.
- Empty images of San Diego Zoo and some random images with Chimps in it.

How to collect such data?

- Images of Chimps and their activities are collected through Google Images API.
- Empty images of Zoo is captured using cameras fixed in Zoo and random images with chimps in it can be generated using any image editor tools.

How big data analytics can uncover the unexpected in your data?

The main challenge in monitoring the activities of wildlife is to eradicate empty images captured in camera. We use Big Data Analytics so that we train system to filter empty images using collected data and Image Classification tool. Now to monitor the activities and location of chimps, we make use of Big Data tools to classify filtered content comparing with collected random images of chimps and its activities.

How does the learning improve your system?

The learning helps to easily monitor data by comparing with small amount of collected data set. By removing empty images it makes system efficient in terms of speed and memory.

Draw the workflow of the proposed system and explain the process of the system.

- The images of Chimps and its activities in different orientations are collected through Google Images API. Empty images of Zoo are collected using the cameras fixed in Zoo and create random images of Chimps in it using any Image editor tools.
- Now collect recorded data from camera and compare it with collected data. This helps us to eradicate empty images from images with Chimps in it.
- Bar code the chimps which helps us to identify the chimps in filtered content. Then use any Big Data tools to classify filtered content based on their location, frequency and activities comparing with the collected random images.
- Store the results into database which can be accessed by the primatologists.

Reference: https://www.cs.ucsb.edu/sites/cs.ucsb.edu/files/docs/reports/tr.pdf