ZUHA AHMAD

DS-670: Capstone: Big Data & Business Analytics

Dr. Jaume

08 April 2017

**DISCUSSION**

The goal of my competitor’s article was to explore a methodology for reliably predicting the location, time, and/or likelihood of future residential burglary in a year’s time. They had several classifiers/algorithms to choose such as SVM, Decision Trees, Neural Networks, Naïve Bayes, and 1NN. We both had wanted to tackle crime forecasting in the United States. However, our focus is different. We are focusing on one algorithm, K means-clustering and decided to build an interactive application in R Shiny from R. The deployment and implementation of visual application for crime forecasting is beneficial and an improvement from our competitors. The major advantage is to control the input parameter with this set of algorithm (K-means clustering). Specifically, we are targeting accidents, burglary, and shootings from year 2012 and year 2016. However, we have more variables in our data that we can also explore in future projects. Time was our independent variable, therefore other types of crimes that were also part of the data were assault calls, disturbance calls, bombing materials, chemical discharge, school disturbances, and stabbing. We had three R scripts to run the application which were global, server, and user interface. For the application to work, we had to install several library packages in our global script. We had to install:

* library(dplyr)
* library(RColorBrewer)
* library(tidyr)
* library(DT)
* library(googleVis)
* library(ggmap)
* library(ggplot2)
* library(shinydashboard)
* library(shiny)
* library(leaflet)
* library(shinythemes)
* library(readr)

Each dataset has a single type of crime with information. Therefore, all of the datasets were merged into one immense dataset called ‘CrimeData1.csv’. For our UI (User Interface) script, we had constructed information to building a widget into R Shiny. Here is our process of our UI code:

dashboardPage(

dashboardHeader(title = "Next 911 Call"),

dashboardSidebar(

sidebarUserPanel("Zuha Ahmad",

image = imgURL),

sidebarMenu(

menuItem("Map", tabName = "map", icon = icon("map")),

menuItem("Data", tabName = "data", icon = icon("database")),

menuItem("Summary", tabName = "summary", icon = icon("analysis"))),

dateRangeInput(inputId="date\_range",

label=" Date Range:",

min = as.Date("2012-03-01"), max = "2016-12-31", end = NULL,

format = "yyyy-mm-dd",

startview = "2012-03-01",

weekstart = 0,

language = "en", width = NULL),

br(),

selectInput("yaxis", "Choose Type Of Crime",

choices = c("Accident Priority" = "ACCIDENT.P",

"Accident" = "ACCIDENT",

"Disturbance Call Priority" = "DIST.P",

"Disturbance Call" = "DIST",

"Assault Call" = "ASSAULT",

"Assault Call Priority" = "ASSAULT.P",

"Assistance Required" = "ASSIST",

"Assistance Required Priority" = "ASSIST.P",

"Bombing Material Found" = "BOMB",

"Burglary" = "BURG",

"Burglary High Priority" = "BURG.HP",

"Burglary Priority" = "BURG.P",

"Chemical Discharge Reported" = "CHEM",

"Disturbance In School" = "SCH\_DIST",

"School Disturbance Priority" = "SCH\_DIST.P",

"Shot Fired" = "SHOT",

"Shooting Priority" = "SHOT.P",

"Stabbing Cold" = "STAB"

),selected = "STAB"

)

),

The map is shown as:

tabItems(

tabItem(tabName ="map",

leafletOutput("mymap",width="100%",height="1000px")

# absolutePanel(id = "controls", class = "panel panel-default", fixed = TRUE,

# draggable = TRUE, top = 60, left = "auto", right = 20, bottom = "auto",

# width = 330, height = "auto")

),

The cluster is shown as:

tabItem(tabName ="Cluster",

plotOutput("clus", width = "100%", height = "800px", click = NULL,

dblclick = NULL, hover = NULL, hoverDelay = NULL,

hoverDelayType = NULL, brush = NULL, clickId = NULL, hoverId = NULL,

inline = FALSE)),

The map is shown as:

tabItem(tabName = "data",

fluidPage(theme = shinytheme("spacelab"),

fluidRow(

column("NATIONAL INSTITUTE OF JUSTICE DATABASE",dataTableOutput(outputId = "table"),width=12),

column(10,p(textOutput("para")))))),

Finally, the summary is shown as:

tabItem(tabName ="summary",

fluidPage(

mainPanel(

plotOutput('plot1'))))

)

)

)

)

The server had several codes, but essentially they link the global and the user interface together.

For accidents in 2012, there is a concentration in crime between the Pearl and University District. In 2016, accidents are heavily concentrated in Pearl, Lloyd, and University district. For both 2012 and 2016, accidents occurred at the bridges and near the river. It also takes place in Government Island which is off-limits to the public. For burglaries in 2012, there are small clusters which surround the Sunset Highway. Though there are 2 orange clusters that were 100+ crime incidents in 2012, in 2016 it turned to 5 orange clusters with 100-400 crime incidents in the Portland region. In addition, there were 2 more orange clusters, one further from Lloyd district and one from Central East Side. The Willamette River that divides Portland from several bridges has heavy amounts of burglary and many territories down to Oregon City. These clusters can refer to the density-based spatial clustering because there are clusters that are closely packed together. Burglary remains to be a common practice of crime and the numbers are abundant in 2012 and 2016. The shots fired (shootings) had increased from 2012 to 2016. The shootings in 2016 had occurred in the Columbia River moving up north to Vancouver towards the Sauvie Island Wildlife Area. The results have indicated that crimes in 2016 with these three variables have increased from 2012 which reportedly was the year of heavy crime. Based on the map, the heavy impact of crimes occurred between river borders of Oregon and Washington which is the Willamette River.

**CONCLUSION**

Though the variables have multiple crime incidents, the heavy concentration is in the metropolis areas of Pearl, University, and Lloyd Districts. This interactive map is great for law enforcements to understand the pattern recognition, data analysis, and image processing. We can identify city from observation databases. As for Portland, we can take it further and learn hotspots based on groups of houses and commercial properties according to its house type, value, and geographic location. Clustering is simple, yet robust to comprehend. It is data mining function which means our analysis had served as a tool to gain insight into the distribution of data to observe characteristics of each cluster.

As we move forward in society, we can make improvements in Internet of Things (IoT) and make smart city initiatives. Community awareness and initiatives in cooperation the Portland community and police departments would be a great improvement. Building smart cities with crime insurance plans will ensure that U.S. citizens (Oregonians) will have enough facilities to lead safely lives. Crime insurance should be designed meticulously, therefore Oregonians would be covered from loss expenditures. Overall, based on our results, we recommend proper cooperation and strategies between law enforcement from incoming and outgoing region of state of north Oregon and the state of south Washington.