Building A Model That Helps Locating Displaced People

Tom Lever

06/08/2023

In this project, we build a model that would help us locate people displaced by the earthquake in Haiti in 2010. More specifically, we build in a timely manner an accurate model that classifies pixels in geo-referenced aerial images of Haiti in 2010 as depicting blue tarps or depicting objects that are not blue tarps. People whose homes were destroyed by the earthquake often created temporary shelters using blue tarps. Blue tarps were good indicators of where displaced people lived.

Our training data was collected likely by applying a Region Of Interest (ROI) Tool to a high-resolution, orthorectified / geo-referenced image of Haiti in 2010. One ROI tool is described at https://www.l3harrisgeospatial.com/docs/regionofinteresttool.html. Classes may be assigned to pixels by defining Regions Of Interest.

Our training data frame consists of $63,241$ observations. Each observation consists of a class in the set $\{Vegetation, Soil, Rooftop, Various\ Non-Tarp, Blue\ Tarp\}$ and a pixel. A pixel is a colored dot. A pixel is represented by a tuple of intensities of color Red, $Green$, and $Blue$ in the range 0 to 255.

According to https://www.esri.com/about/newsroom/insider/what-is-orthorectified-imagery/, an orthorectified image is an accurately georeferenced image that has been processed so that all pixels are in an accurate (x, y) position on the ground. Orthorectified images have been processed to apply corrections for optical distortions from the sensor system, and apparent changes in the position of ground objects caused by the perspective of the sensor view angle and ground terrain.

We load a data frame of classes and pixels based on an orthorectified image of Haiti at https://www.kaggle.com/datasets/billbasener/pixel-values-from-images-over-haiti?datasetId=1899167.

```
data_frame_of_classes_and_pixels <- read.csv(
    file = "Data_Frame_Of_Classes_And_Pixels.csv"
)
head(x = data_frame_of_classes_and_pixels, n = 2)
```

```
#        Class Red Green Blue
# 1 Vegetation  64    67   50
# 2 Vegetation  64    67   50
```

```
tail(x = data_frame_of_classes_and_pixels, n = 2)
```

```
#              Class Red Green Blue
# 63240 Blue Tarp 132   139  149
# 63241 Blue Tarp 133   141  153
```

We build binary classifiers that classify pixels as depicting blue tarps or depicting objects that are not blue tarps. We may ignore non-binary classifiers that predict probabilities for all classes and may be used to locate pixels that more likely depict blue tarps than objects that are not blue tarps. We may ignore non-binary

classifiers since the intensity space for pixels representing blue tarps is distinct from the intensity space for pixels representing objects that are not blue tarps. See "Figure 1: Distribution Of Classes In Intensity Space".

```r
library(rgl)
vector_of_possible_colors <- c("green", "brown", "black", "orange", "blue")
number_of_observations <- nrow(data_frame_of_classes_and_pixels)
column_of_numerical_representations <- rep(0, number_of_observations)
condition <- data_frame_of_classes_and_pixels$Class == "Vegetation"
column_of_numerical_representations[condition] <- 1
condition <- data_frame_of_classes_and_pixels$Class == "Soil"
column_of_numerical_representations[condition] <- 2
condition <- data_frame_of_classes_and_pixels$Class == "Rooftop"
column_of_numerical_representations[condition] <- 3
condition <- data_frame_of_classes_and_pixels$Class == "Various Non-Tarp"
column_of_numerical_representations[condition] <- 4
condition <- data_frame_of_classes_and_pixels$Class == "Blue Tarp"
column_of_numerical_representations[condition] <- 5
vector_of_colors <- vector_of_possible_colors[column_of_numerical_representations]
plot3d(
    x = data_frame_of_classes_and_pixels$Red,
    y = data_frame_of_classes_and_pixels$Green,
    z = data_frame_of_classes_and_pixels$Blue,
    col = vector_of_colors,
    xlab = "Red",
    ylab = "Green",
    zlab = "Blue"
)
rglwidget()
```

In order to build binary classifiers, we create a data frame with a column of indicators of whether of not a pixel depicts a blue tarp instead of a column of classes.

```r
library(rgl)
number_of_observations <- nrow(data_frame_of_classes_and_pixels)
column_of_indicators <- rep(0, number_of_observations)
condition <- data_frame_of_classes_and_pixels$Class == "Blue Tarp"
column_of_indicators[condition] <- 1
factor_of_indicators <- factor(column_of_indicators)
data_frame_of_indicators_and_pixels <- data.frame(
    Indicator = factor_of_indicators,
    Red = data_frame_of_classes_and_pixels$Red,
    Green = data_frame_of_classes_and_pixels$Green,
    Blue = data_frame_of_classes_and_pixels$Blue
)
vector_of_random_indices <- sample(1:number_of_observations)
data_frame_of_indicators_and_pixels <- data_frame_of_indicators_and_pixels[vector_of_random_indices, ]
head(x = data_frame_of_indicators_and_pixels, n = 2)
```

```
#       Indicator Red Green Blue
# 56484         0 178   160  116
# 6477          0  79    74   65
```

```
tail(x = data_frame_of_indicators_and_pixels, n = 2)
```

```
#       Indicator Red Green Blue
# 53098         0 171   166  152
# 17872         0  94    90   69
```

We use 10-fold cross-validation to evaluate the performance of 5 classifiers. A classifier will classify a pixel as depicting a blue tarp or depicting an object that is not a blue tarp.

We use 10-fold cross-validation to evaluate the performance of logistic-regression models.

```
library(TomLeversRPackage)
plot_ROC_curve_for_cross_validated_models(
 type_of_model = "Logistic Regression",
 formula = Indicator ~ Red + Green + Blue,
 data_frame = data_frame_of_indicators_and_pixels,
 number_of_folds = 1,
 step = 0.2
)
```

```
# Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred


#
# vector_of_predicted_response_values     0     1
#                                   1 61219  2022


# Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred


#
# vector_of_predicted_response_values     0     1
#                                   0 60810   126
#                                   1   409  1896


# Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred


#
# vector_of_predicted_response_values     0     1
#                                   0 61141   211
#                                   1    78  1811


# Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred


#
# vector_of_predicted_response_values     0     1
#                                   0 61163   251
#                                   1    56  1771


# Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```
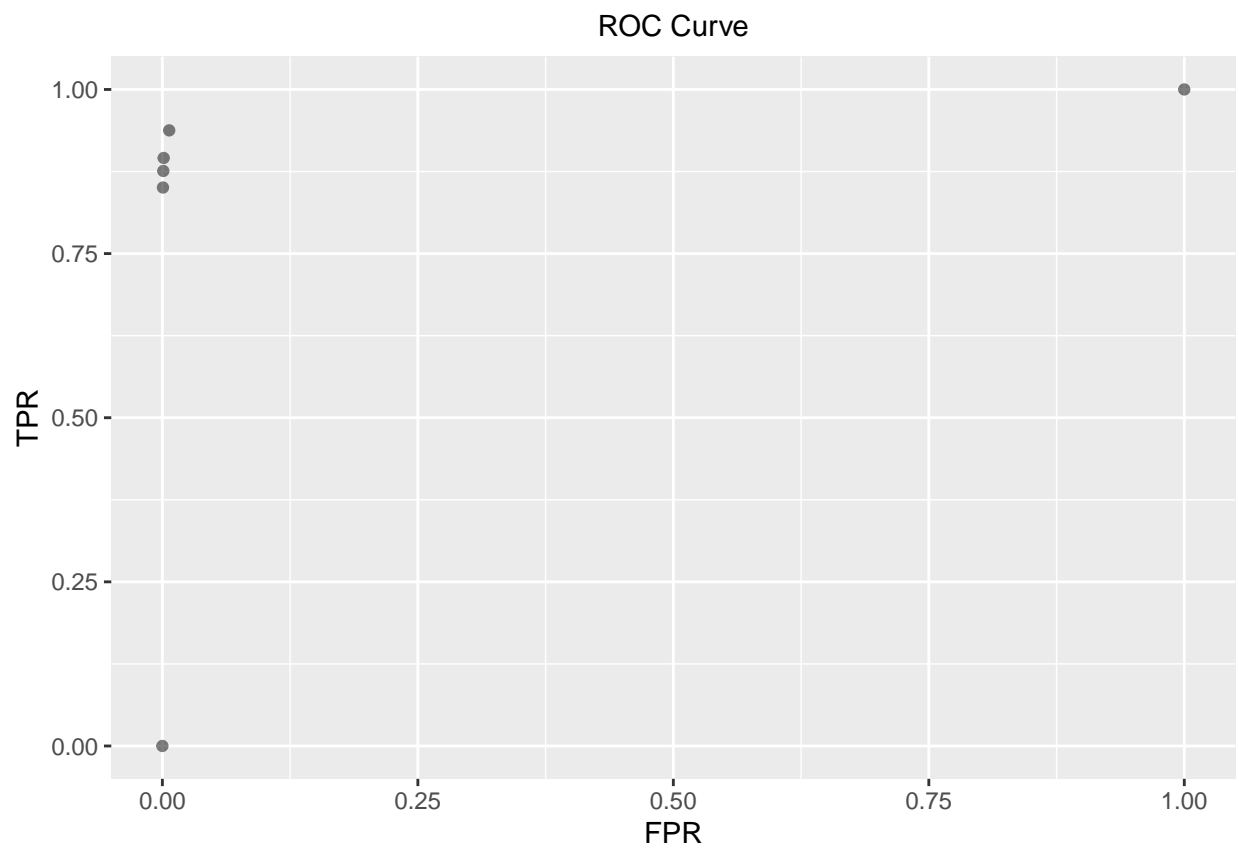
```
#
# vector_of_predicted_response_values      0      1
#                                    0 61183    302
#                                    1    36   1720


# Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred


#
# vector_of_predicted_response_values      0      1
#                                    0 61219   2022
```

ROC Curve

TODO: Consider 5 number summaries and means and standard deviations of Red, Green, and Blue. and across classes.