An Analysis of the Driving Habits of Taxi Drivers in Midtown Manhattan Based on Publicly Available Webcam Data

Brian Luczkiewicz
15-394 Data Privacy
Assignment 6, second hand-in

Abstract

Publicly accessible webcams potentially provide a 24/7 stream of data for use in the pursuit of research. However, efforts to use this data have been limited thus far and it is not yet known whether or not the data is of sufficient quality or accessibility to be very useful. In this paper, a webcam observing a midtown Manhattan intersection is used to draw conclusions about the behavior of taxicab drivers.

Introduction

Much of the uncertainty regarding the applicability to webcam data in research boils down to a few problems. These include the low resolution and lens-quality of most webcams, the carelessness with which they are placed, and the complex flash and java-based delivery mechanisms. Additionally, some webcams allow site visitors to control the camera remotely, which makes it more difficult to use the images in research.

This paper concentrates on a single public webcam that overcomes many of these complaints. The webcam [1] delivers a 703x575 pixel image in jpeg format every five minutes, along with a 352x288 pixel image every 20 seconds. The camera provides a particularly good view at the front line of the southbound traffic on 5th Ave. at 45th st. We will focus on the larger image from the webcam. Every image generated by the webcam over the one-week period between March 30^{th,} and April 6th, 2006.

New York taxi-drivers are famous for their aggressive driving habits. Via casual observation we can see that taxi-drivers tend to position themselves as closely as possible to the stopping line on when traffic signals demand a stop, presumably to get the fastest possible start when the light turns green. The data set from this webcam provides an ideal set of images to confirm or refute this observation.

Methods

Using an automated script, 2005 images were sampled from the camera at approximately five-minute intervals over a one-week period. Initially, it was thought that the images would yield easily to automated analysis, but due to a variety of factors this was impractical. As a result, the images were scrutinized individually, by hand. Five pieces of information were collected from each image:

Figure 1: Depiction of Scrutinized Area

- Number of southbound vehicles visible partially or fully in the scrutinized area.
- Number of southbound visible partially or fully in the scrutinized area
- Whether or not traffic is "stopped"
- If traffic is "stopped", how many vehicles make up the "front line" of traffic
- If traffic is "stopped", how many cabs are part of the "front line" of traffic

Figure 2: Sample webcam images depicting stopped traffic

Unfortunately, we have no view of the traffic light in the image. Thus, a number of factors are taken into account in determining if traffic is "stopped" in a given image. Not all images in which there are cars located directly behind the white line represents stopped traffic. It is assumed that if there are any westbound vehicles on 45th st, then the

traffic on 5th avenue is stopped. It is also assumed that if a car exhibits a blurring parallel to the direction of 5th Ave, then it is in motion. If there are no cars or no cars close to the white line and stopped, then traffic is not "stopped".

The "front line" of traffic is made up of the cars positioned the farthest south in each of the five possible lanes of traffic such that at least part of the car is in front of the green line.

Results

Out of the 2016 images sampled, 672 of them depicted stopped traffic. Of these, 563 of them included cabs. In order to gauge the hypothesis of this paper, we'll examine the ratio between cabs/vehicles and front line cabs/front line vehicles in these 512 data entries.

$$CabAggressiveness = \frac{\left(\frac{FrontLineCabs}{FrontLineVehicles}\right)}{\left(\frac{Cabs}{Vehicles}\right)}$$

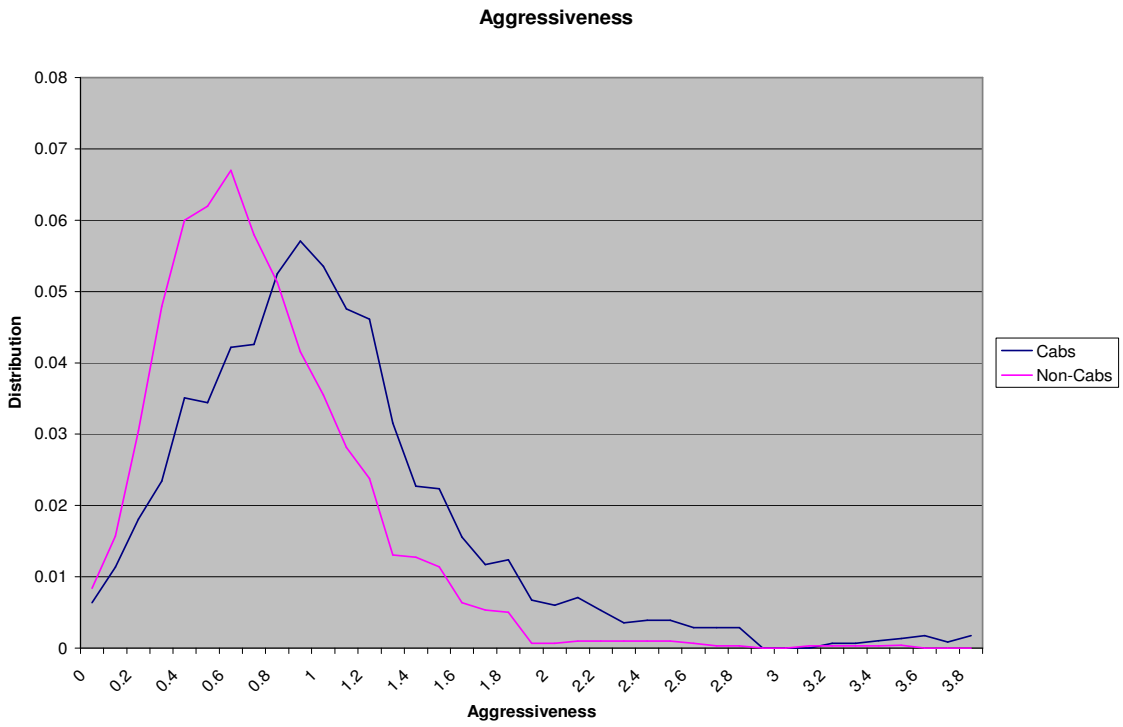
Of the images sampled, 596 of them depicted stopped traffic and included at least one non-cab vehicle. We can construct a similar ratio for non-cabs, as follows:

$$NonCabAggressiveness = \frac{\left(\frac{FrontLineNonCabs}{FrontLineVehicles}\right)}{\left(\frac{NonCabs}{Vehicles}\right)}$$

Applying this equation to the collected data yields two new columns which records, by our loose definition, the aggressiveness of cab drivers in that image normalized around 1.0. A t-test analysis of the data with alpha=0.05 yields the following:

	Cab Aggressiveness	***Non-Cab Aggressiveness***
Mean	1.125338475	0.932983548
Variance	0.392382529	0.221250805
Observations	563	596
t Stat	5.88548133	
P(T<=t) one-tail	2.67365E-09	
t Critical one-tail	1.646318694	
P(T<=t) two-tail	5.3473E-09	
t Critical two-tail	1.962245385	

This shows that the difference between the test group (Cabs) and the control group (Non-Cabs) is statistically significant. A graph of the distribution of the results is also provided:



Discussion

The data and subsequent analysis has shown that given our operational definitions, cab drivers at this intersection are more aggressive than other drivers to a statistically significant degree.

It is important and interesting to note that although the cab drivers in this study were almost entirely anonymous, they were observed without their knowledge via a publicly accessible webcam. This could motivate relevant discussion as to whether the privacy of the drivers was violated.

Lastly, it should be noted that the data collected to this research could possibly be applied towards answering a wide variety of questions about driving and this intersection. There is certainly sufficient information to track traffic volume in both cabs and vehicles in general, and since all of the data is time stamped, it could be correlated with weather data, local event schedules, or traffic light patterns. The potential applications of this data set have certainly not been exhausted in this study.

References

[1] http://riotmanhattan.com/riotmanhattan/index.php?page=webcam