Distracted Driving Data Analysis Proposal

Joseph Blubaugh

02 October 2016

Data

The data in this project are of 8 driving simulations for 66 individuals ranging from 3,000 to 30,000 observations per simulation. There are over 6.7 million observations in the entire dataset. The data from each simulation includes likelihood scores for 8 facial expressions recorded at a fixed interval of .03 seconds. Stimuli data which records targetted events that were introduced into each simulation and basic demographic data on each subject are also available.

Table 1: Combined Simulation and Stimuli Datasets

ID	Age	Gender	Event	Time	Anger	Contempt	Disgust	Fear	Joy	Sad	Surprise
T001-001	Y	M	No Event	0	0.01	0.022	0.004	0.054	0.526	0.096	0.001
T001-001	Y	M	No Event	0.033	0.01	0.022	0.004	0.054	0.526	0.096	0.001
T001-001	Y	M	No Event	0.067	0.01	0.022	0.004	0.054	0.526	0.096	0.001
T001-001	Y	M	No Event	0.1	0.008	0.019	0.003	0.038	0.535	0.105	0.001
T001-001	Y	M	No Event	0.133	0.009	0.038	0.016	0.004	0.69	0.018	0
T001-001	Y	M	No Event	0.167	0.01	0.045	0.014	0.003	0.716	0.016	0

There were several misspelled descriptions in the stimuli files so I reconstructed the event names. At least one of the events appears to be miscoded. There is only one simulation in trial 006 that has a Mathematical Question event. This appears to be miscoded and should probably be coded as Emotional Question instead.

Table 2: Count of Simulations by Event Type

Event	001	002	003	004	005	006	007	008
No Event	63	60	61	62	62	60	59	62
Analytical Questions	0	0	0	0	62	0	0	0
Emotional Questions	0	0	0	0	0	59	0	0
Failure Event	0	0	0	0	0	0	0	59
Mathematical Questions	0	0	0	0	62	1	0	0
Texting	0	0	0	0	0	0	58	0
Texting and Talking	0	0	0	0	0	0	0	29

The published literature on this data showed evidence that texting and driving contributes significantly to poor driving performance. Driving performance data is unavailable in this analysis.

Analysis Proposal

The plot below shows the output of two driving simulations for a single subject (T086), one with no events (Sim 004), and one with texting events (Sim 007). While we do not know the subject's driving performance in any of the simulations, it is clear to see that there are changes in the subject's facial expressions during the texting events compared to the driving simulation where no events occur.

I propose that this analysis focus on building a model to detect whether the subject is texting or otherwise distracted based on demographics and changes in facial expressions. I would also like to leave in the time series component and experiment with complex models such as neural nets or support vector machines. I am particularly interested in measuring the similarities in facial expression changes based on encountered events. I would also like to understand if one model performs reasonably well on all or most subjects, or are the subjects so unique that only individual models would perform well.