

Distracted Driving Initial Analysis

Joseph Blubaugh

17 October 2016

Overview

This analysis is focussed on trying to detect texting events from subjects participating in the driving simulation trial 007. In the distracted driving experiment, trial 004 was used as a practice run where the driver drove approximately 11 kilometers on a crowded four lane road with a speed limit of 70 kph. There was a period in the drive where the driver had to take a short detour, but otherwise the drive was a straight line. No events occurred in this simulation and so trial 004 can be used as a baseline to compare against trial 007 where the driver encounters texting events on the same road and in the same conditions.

Establishing a Baseline

In order to establish a baseline, the overall mean was computed for each subject's emotional likelihood over the entire simulation. The mean was subtracted from each observation in trial 007 to calculate a baseline starting point.

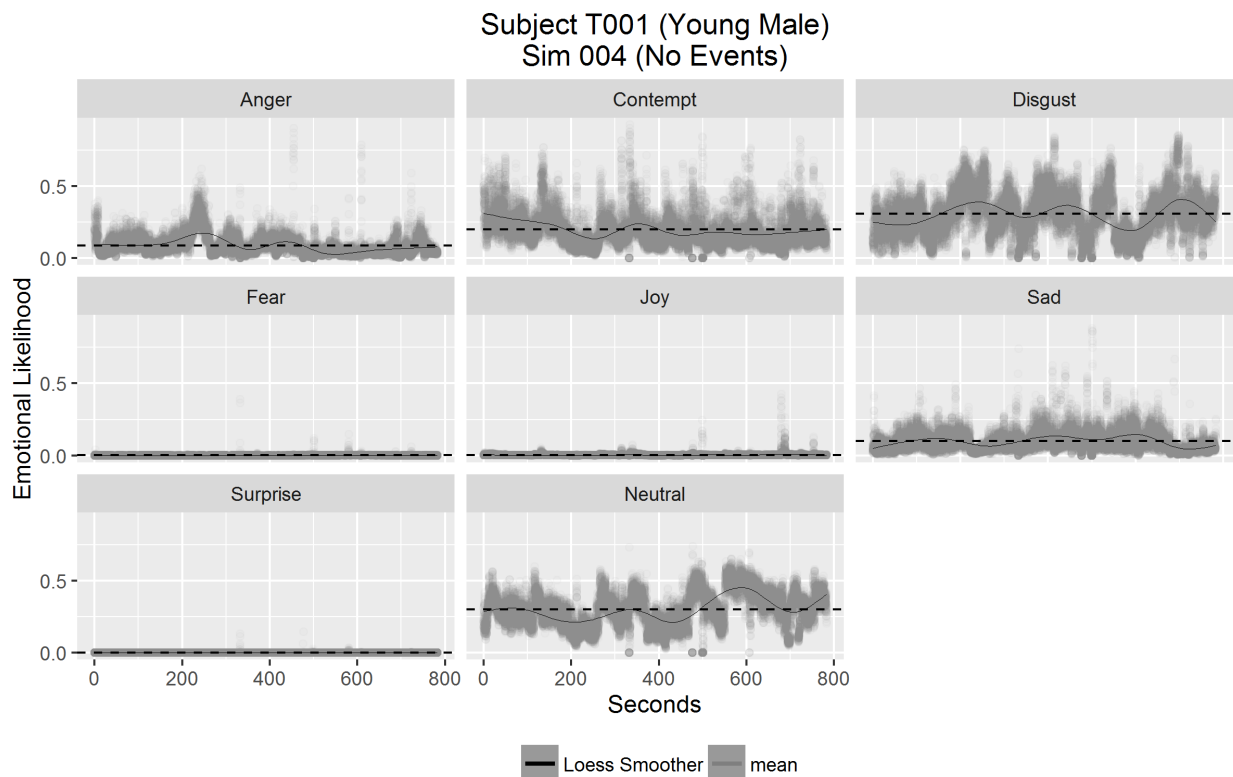


Figure 1:

Every observation of trial 004 and 007 are plotted below. Zero on the y-axis represents the baseline value for the Emotional Likelihood in trial 004. Gray points represent no events in the drive and yellow points represent texting events during the drive which only occur in trial 007. The black and red lines are loess smoothers to show the overall trend. This particular subject shows both positive and negative values of neutral, disgust, and contempt vs the baseline indicating that any deviation from the baseline regardless of direction may indicate texting. The disgust emotion appears to show more variation during texting events relative to non-events. There are also significant shifts away from the baseline for non-events. The most noticeable emotion is anger where during a non-event, anger increases rapidly vs the baseline and remains high even after a texting event occurs. This example may support the claim in the published paper that residual effects of texting seem to linger in driving behavior even after a texting event ends. These types of sudden shifts in emotional likelihood could make it difficult to detect texting events based on facial expressions alone.

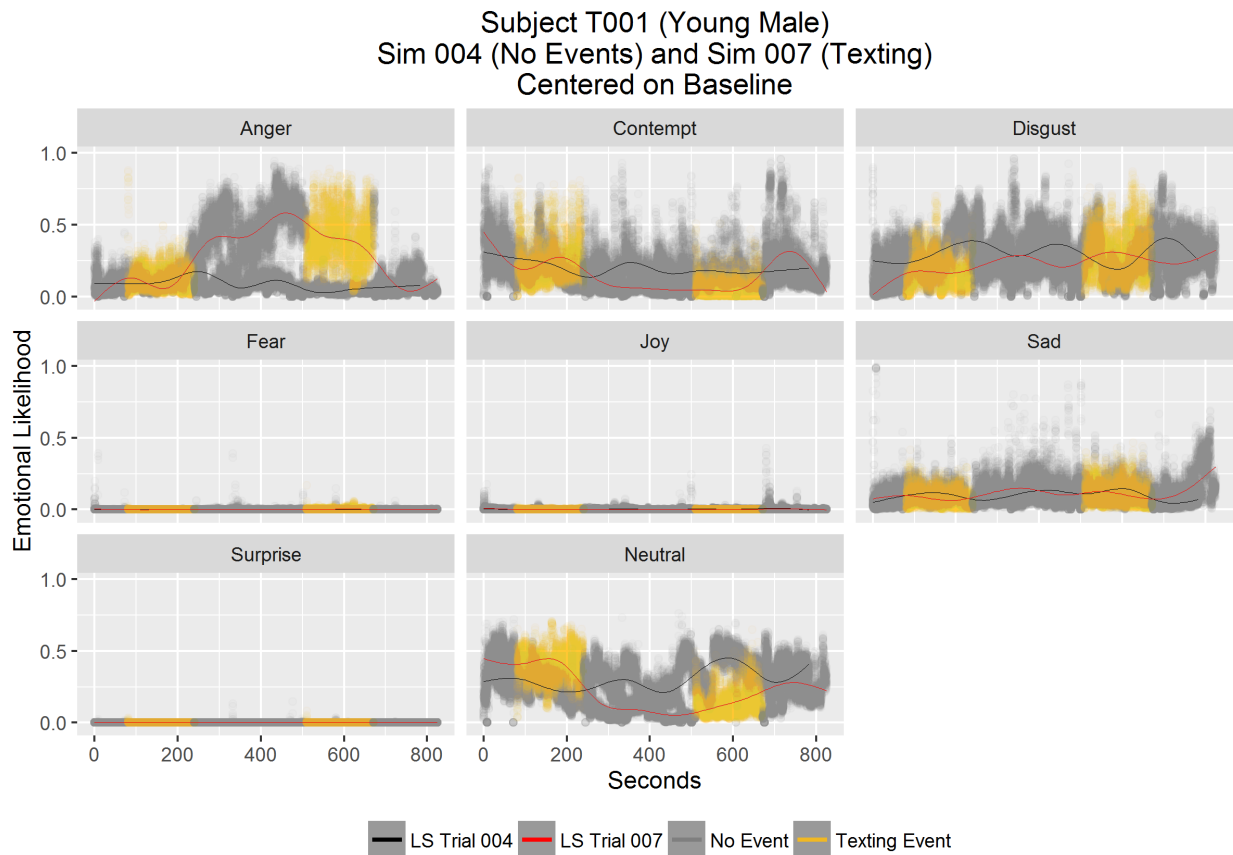


Figure 2:

The plot below shows individual loess smoother plots for each individual that participated in the non-event trial (004) and texting trial (007). The main point of this plot is to show differences in variation between the two trials. The lines are significantly more “wiggly” in the bottom row (007) than the top row (004) for all 5 emotions. Fear, joy, and surprise had smaller scale differences and were removed from this view to give more plotting space

to the other emotions.

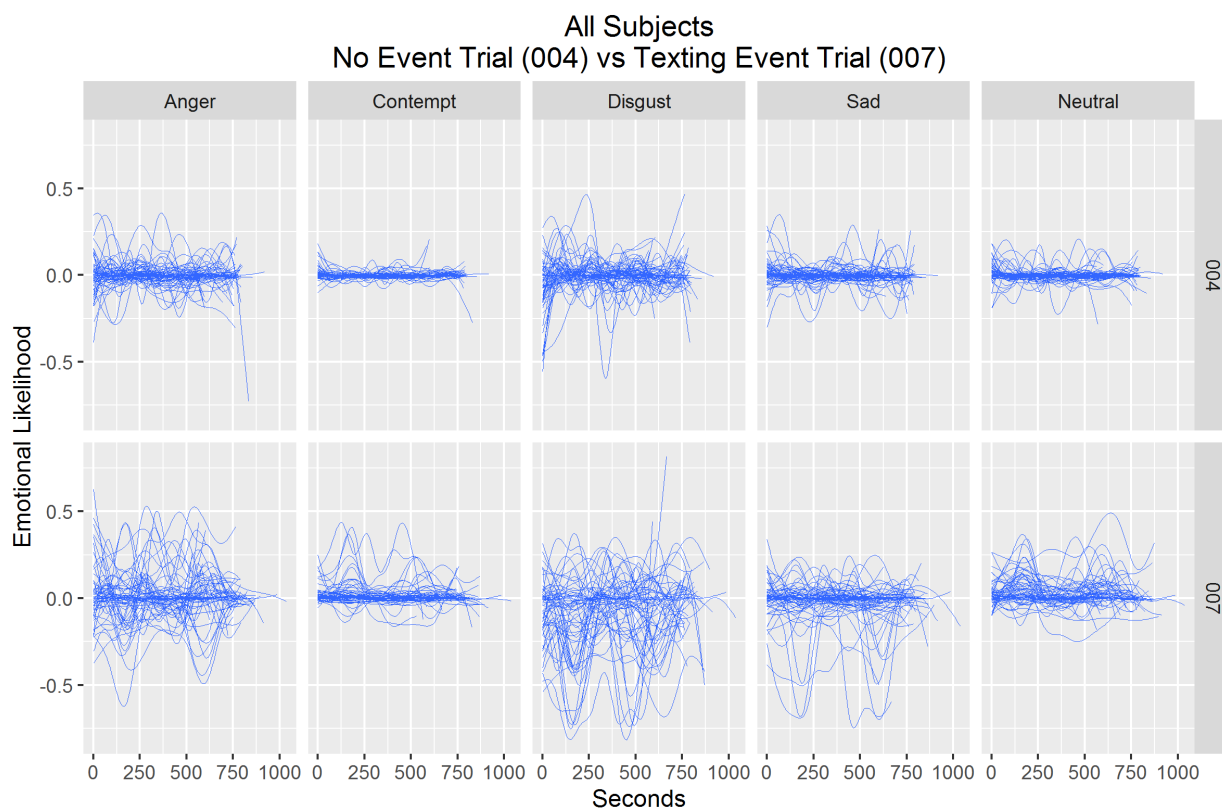


Figure 3:

Exploring a Macro Level Model

One of the interesting aspects of this analysis is to explore the differences between subject's reactions. In this case we are limited to facial expression data and simple demographics so its worth investigating whether a single model can effectively classify a texting event. Since the baseline has been established and the mean effect has been removed, I only used trial 007 in this portion of the analysis. In each of the texting trials, texting events occur twice with periods of non-events in between.

In 2012 Ford sponsored a competition to develop a model that could successfully detect whether a driver was paying attention to the road. The data was similar to this project and the winner of the Ford competition wrote that instead of approaching the analysis as a time series problem, he instead averaged all values over the entire trial so that he had a single observation per trial. I have elected to try a similar approach first. I computed the min, max, mean, var, and iqr for every simulation, grouped by the event. For all simulations, 119 observations and 40 descriptive statistics were available.

The 119 observations were split into a training and testing set with a 60/40 ratio. A neural net and random forest model were both applied to the training set with default parameters. An overall quality measure of $(\text{Sensitivity} + \text{Specificity}) / 2$ was used to evaluate each model.

Neural Network Confusion Matrix

	Predicted	
Actual	0	1
0	34	0
1	0	37

(Training Set) Overall Performance: 1

Random Forest Confusion Matrix

	Predicted	
Actual	0	1
0	34	0
1	0	37

(Training Set) Overall Performance: 1

Neural Network Confusion Matrix

	predicted	
actual	0	1
0	15	11
1	7	15

(Testing Set) Neural Net Overall Performance 0.6293706

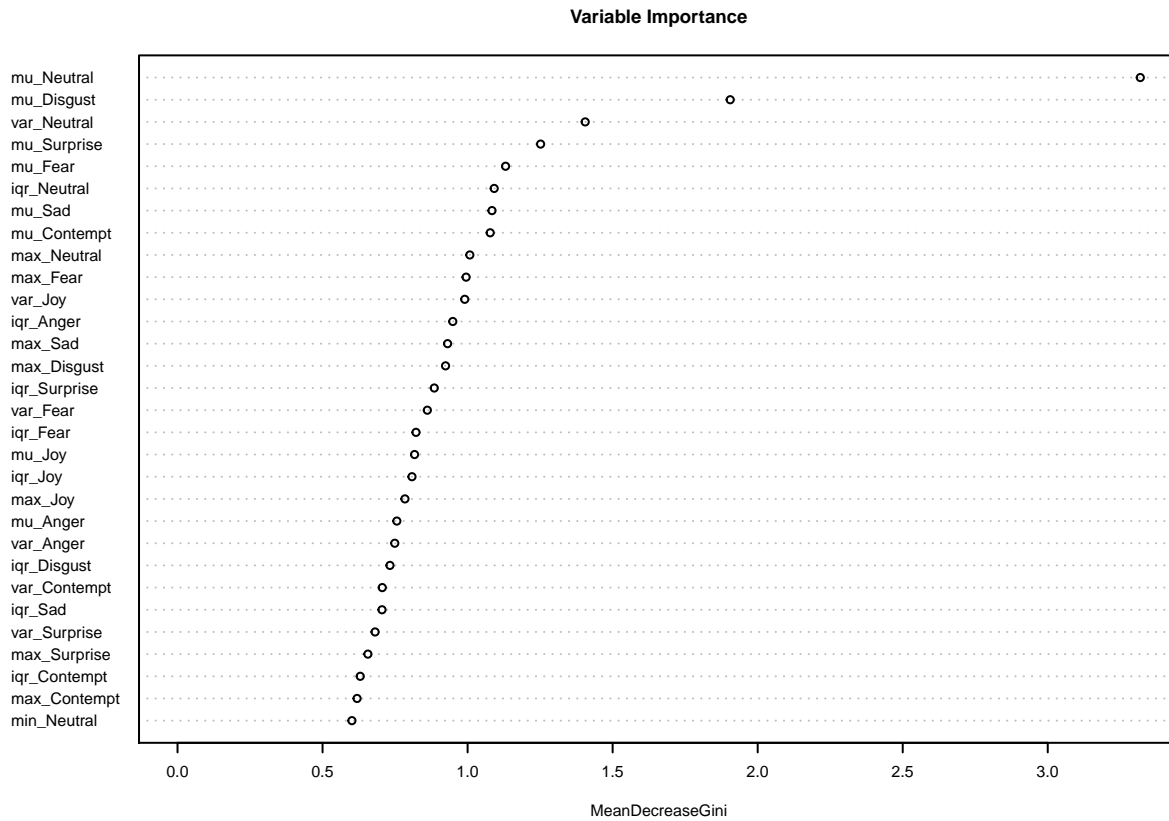
Random Forest Confusion Matrix

	predicted	
actual	0	1
0	12	14
1	9	13

(Testing Set) Random Forest 0.5262238

The training set performance is near perfect in most cases, but the testing set performance is comparable to a coin flip. It is obvious from the results of the confusion matrix that both models are guilty of overfitting. I tried decreasing the amount of data considered by the models and got similar results. Some minor experimenting with tuning parameters also had no effect. An inspection of the variable importance plot of the random forest model shows that the mean of the neutral emotion tends to have the highest impact on texting prediction. This in itself might be a reason why the models struggle to fit this type of data. The

neutral emotion might be the best emotion to predict texting simply because people are unique tend to react differently when its comes to texting events or stimuli in general.



Conclusions and Next Steps

The summary statistics calculated at each event does not provide a good enough foundation to adequately detect event changes. Furthermore, it seems there is evidence that a one size fits all model will not work for this problem given that the neutral face seems most important when it comes to predicting texting. The restrictive size of the aggregated dataset makes overfitting a challenge and in fact may filter out the signals needed to detect texting events. In the next phase of this analysis, I would like to focus on looking at the same trials in the original times series format. The nature of the experimental design affords the ability to create a training and testing set for each subject in the texting trial. I will also spend a significant amount of time experimenting with model tuning.