CS 4641 Project Proposal

Alexander Faché[1], Roshni Ghosh[2], Hisham Temmar[3], Rushabh Varia[4], and Suraj Vuduta[5]

[1]afache3@gatech.edu
[2]rghosh40@gatech.edu
[3]htemmar@gatech.edu
[4]svuduta3@gatech.edu
[5]rvaria3@gatech.edu

February 18, 2020

1 Introduction

We will be analyzing US Traffic Accidents and feeding in additional information with US Weather Events in order to determine locations where driving conditions are most susceptible to weather factors.

There are strong relationships between road safety and weather conditions, as extreme conditions can create safety hazards along roads which leads to higher crash rates. The US Department of Transportation Road Weather Management Program reports that annual averages from 2007-2016 show 15% of vehicle crashes occurred due to wet pavement, 10% due to rain, 4% due to snow, and 3% due to ice ("How Do Weather Events Impact Roads?", 2018). It is paramount that drivers are aware of the most dangerous locations in their commutes during extreme weather, so they can take extra caution when passing or avoid those areas entirely. Thus, we aim to develop a risk assessment for regions that drivers can use to evaluate driving conditions and take extra precautions.

2 Methods

Our first priority will be to identify which features we can exclude from our model. Our first data set contains 49 features per sample, so a fair amount of dimensionality reduction will be required. For instance, Street/City/County/State features could be excluded in favor of latitude/longitude instead, thereby grouping regions by proximity rather than text based descriptors. We will initially aim to model the Atlanta area. If successful, we will then aim to generalize our model to more and larger regions, with a larger variety of driving conditions.

This dimensionality reduction will likely be performed with the help of principal component analysis (PCA) (Mercer, 2011). Once the key features have been extracted, we will proceed with developing a model such that a user can input their location, weather forecast (and other extracted features) and the model will generate a risk score for the given parameters.

We would like to employ an unsupervised clustering approach to categorize our results once PCA is complete. The generate clusters can then be defined along a risk score (Ferstl, n.d.). The crash report data set also includes a severity score for each crash, which we will tie into our risk assessment model as well. Once produced, these scores can be visualized as a heatmap over the Atlanta region, where changing parameters like weather and other key features will update the heatmap displaying our risk factors.

3 Potential Results and Checkpoint

We expect our model to accurately map weather conditions, locations, and other important features to a accident risk rating. This will help drivers to understand the potential dangers when several factors all line up for a perfect storm. Additionally, the PCA performed in the first stage of the project may help uncover some of the most important factors to consider when making road design decisions and in driver awareness education.

4 Discussion

The optimal outcome for this project would be a model that allows drivers to determine the risk of driving given weather conditions in the their region of driving interest.

With time permitting we will also try gathering current traffic and weather information not present in our data set to determine if this model can be applied outside of 2016-2019 data set range.

References

Ferstl, R. W., Kanzler. (n.d.). Time-hierarchical clustering and visualization of weather forecast ensembles. , 1–10. Retrieved from https://www.in.tum.de/fileadmin/w00bws/cg/Research/Publications/2016/TimeHierarchicalClustering/timeclus_preprint.pdf

How do weather events impact roads? (2018). *Federal Highway Administration*. Retrieved from https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm

Mercer, L., Richman. (2011). Identification of severe weather outbreaks using kernel principal component analysis. , 1–6. Retrieved from https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050911X00054/1-s2.0-S1877050911005084/main.pdf