CSCI 353/795 | Fall 2020

Final Project Proposal

Kamil Sachryn Hannah Do Isaac Lapides

1 Project Title

Our title is “Look Before You Leap: Leveraging Predictive Models to Improve Automotive Safety and Travel Time”

2 Description

2.1 Previous Work

Being able to better predict traffic data for use in route optimization via the use of live social media data could create new and exciting avenues for the field, possibly allowing for unprecedented accuracy for areas with constant slowdowns or accidents.

Minh (2016) attempted to detect traffic events based on Twitter data in Finland. The project focused on using twitter data to find any traffic event and to display it, rather than attempt to find a route free of traffic obstructions.

Furthermore Murueta et al (2019) conducted a study detailing strategies to redirect vehicles around congested areas, allowing for prediction of congested areas and ways to avoid such events.

2.2 Objectives

Our objectives are the following:

A. Replication and development of previous predictive models for automotive traffic and accidents in urban areas
    a. Implementation of various machine learning mechanisms (e.g. SVM, Random Forest, Decision Tree, etc.)
    b. Integration of heterogeneous datasets
    c. Using cross-validation and testing models on different cities to see which are “twinned”
2. Combining the predictive model with a pathfinding algorithm which optimizes for the “safest route.”
    a. Implementation of a pathfinding algorithm
    b. Designing a cost function which appropriately assigns a metric of safety per street-segment-hour
    c. Finding an appropriate weighting threshold between the safety cost function and conventional speed

2.2.1 Challenges

A. Data Mining
    a. Tweet extraction and processing
B. Feature Engineering
    a. Combining several GIS datasets with tweet dataset
    b. Appropriate sampling to compensate for highly imbalanced classes
C. Algorithmic
    a. Implementing specific ML techniques to create and measure predictive models
    b. Implementing an appropriate pathfinding algorithm which can be meaningfully weighted by safety

2.2.2 Innovations

- Addition of tweets as a feature to predictive model
- Creating a pathfinding algorithm that optimizes for safety
- Cross-validation of the model against other cities

2.2.3 Synthetic example

A. Given a specific time frame and location, our machine learning model will produce a heat map similar to the image below provided by Wilson (2018).

The red dots represent the predicted Accident Risk on the map.

B. Given a time frame, an origin, and a destination, we will use our model to produce a set of directions which will minimize the trip's accident risk.

3 Tasks and Roles

Hannah

Using Twitter API to collect location hashtags within a given timeframe, pre-processing the tags to get the frequency of the location corresponding to the location index in the existing table. (optional) may work on the main function and visualizing the accident risk on a heat-map.

Isaac

Implementation of machine learning techniques (SVM, Decision Tree, Random Forest, DNN) and initial model optimization. Implementation of pathfinding algorithm optimizing for safety.

Kamil

Using relevant datasets (Road, Weather, etc) in order to create an optimized route based on current conditions as well as known road conditions via machine learning techniques.

4 Topics

4.1 Natural Language Processing (NLP)

This topic will be demonstrated in pre-processing Twitter hashtags into tokens and counting the frequency in order to add the popular location feature.

4.2 Machine Learning

1. Support Vector Machines (SVM)
2. Decision Tree (DT)
3. Random Forest (RF)
4. Deep Neural Network (DNN)

The four ML models will process the vectors with additional feature sets to predict the possible accident risk.

5 Deliverables

5.1 Timeline

- 10.27.2020 - Week 8
  Collection & pre-processing of the features (weather, road infrastructure, Twitter) and implementation of the ML models

- 11.10.2020 - Week 10
  Evaluation based on the current model and features, making adjustments based on the result + visualization of the result

- 12.1.2020 - Week 13
  Analyze the data and prepare for the final paper and the presentation

5.2 Required Objectives

1. Evaluation of the car accident risk between the predicted data and the existing record in different cities

2. Visualization of the difference between predicted risk and past accident records on each city's map

3. Finding the optimal 'safe' route that would avoid accident prone zones

We plan to submit the finished source code, research paper, and a short video (or a poster) of the project by the end of the semester.

6 Evaluation

We will be using accuracy, precision, recall, f1 score and AUC measure as our evaluation metric as they seem to be commonly used in machine learning algorithms to measure the difference between predicted and actual outcome. In our project, we will be measuring the difference between the predicted probability of car accidents and the actual car accident data that is retrievable from the Countrywide Traffic Accident Dataset by Moosavi et al. (2019). We are expecting that adding custom features such as weather, road infrastructure, and twitter location tags would increase the accuracy in predicting the accident probability.

For the optimal safe route calculation, we will be visualizing the result on the map to see if the fastest optimal route is avoiding accident prone areas.

7 References

Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. (2019). *A Countrywide Traffic Accident Dataset* (arXiv) preprint arXiv:1906.05409

Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. (2019) *Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights.* In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM.

Wilson, D. (2018) *End result: an accident risk heat map* https://medium.com/geoai/using-machine-learning-to-predict-car-accident-risk-4d92c91a7d57

Perez-Murueta, P., Gómez-Espinosa, A., Cardenas, C., & Gonzalez-Mendoza, M. (2019). Deep Learning System for Vehicular Re-Routing and Congestion Avoidance. *Applied Sciences, 9*(13), 2717. doi:10.3390/app9132717

Minh, H. D. (2016). Detection of Traffic Events from Finnish Social Media Data. *Detection of Traffic Events from Finnish Social Media Data*.