Traffic Relations

Extended Abstract*

David Nyberg
CU Boulder
Boulder, Colorado
dany3289@colorado.edu

Eric Ha
CU Boulder
Boulder, Colorado
erha5113@colorado.edu

Nicholas Sugarman
CU Boulder
Boulder, Colorado
shadowelecman@hotmail.com

ABSTRACT

This paper is in ACM SIG format and covers an overview of our data mining project that will be going on during Spring 2018 at University of Colorado, Boulder.

KEYWORDS

ACM, LaTeX, Traffic, CU Boulder

ACM Reference Format:
David Nyberg, Eric Ha, and Nicholas Sugarman. 2018. Traffic Relations: Extended Abstract. In *Proceedings of Traffic Relations Data Mining 2018 (BOULDER, 18).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Through the access of a large traffic data set, we seek to find knowledge involving a correlation between seasonal differences and its application towards traffic volume. An example of a finding would be that there is an increase in traffic during winter seasons. We hope to apply this knowledge towards building more efficient roadways that mitigate traffic that occurs during seasons of high traffic. Interesting knowledge that we also seek to find would be to find a correlation between traffic volume and the type of road in which it occurs. To extend our previous example, we may find that there is a larger traffic volume at rural roads during the winter compared to other types of roads. Another interesting attribute to look at along with the previous could be lanes of traffic, depending on how many lanes a road may have, does this impact how many people drive on this road, and when do they do it? All these are important questions to any commuter or city planner or even construction companies trying to repair and build roads.

2 PREVIOUS WORK

Traffic issues are an ongoing problem in the United States and lots of previous work has been done to research and try to fix traffic around the world.

https://www.zmescience.com/research/technology/google-maps-traffic-05443/

This article talks about Google maps and how it uses predictive technology in tandem with traffic data to predict when and where traffic jams will occur. It can predict future traffic jams by referencing past traffic data and comparing it with current conditions to see if there is a significant overlap. Similar work has been done to determine what days are more dangerous to drive on then others

https://www.bactrack.com/blogs/expert-center/35042821-the-most-dangerous-times-on-the-road.

This data is used to measure when drunk drivers are most likely to be on the road. https://www.nature.com/articles/srep37300

Research has been done which investigates what weather conditions are most likely to put people in danger and where. The study also tries to take factors such as socioeconomic status into account, which is a little beyond the scope of our project but still fascinating.

http://kalw.org/post/driving-apps-waze-are-creating-new-traffic-problems

This article writes about how traffic phone applications such as waze which is similar to Google maps showing users the fastest route to a location are actually creating more traffic issues. This is closely related to our project as we are going to look at when people use local roads depending on the time of year and how many lanes they have, this article mentions how apps like waze are sending people onto these small rural roads in order to find a 'faster' path but it is actually slowing down traffic according to their studies.

2.1 Proposed Work

Our data will require cleaning to ensure there are no null values or odd circumstances (sporting event, concert, city evacuation, etc.) which could affect the data results. We will also have to reduce the data to remove any variables that we will not want to study or that will not aid our investigation. This is important to do so we can speed up our analysis times as we have over 7 million data points. We will also have to transform the data to ensure that it is normalized and that the traffic volumes for different areas are on an even scale. We then need to create interesting models that show a correlation between some of our attributes that we can report as an interesting find, this will be the majority of the work. Finally we would want to visualize our data in a meaningful way with Python or another open source tool. Our study is similar to Google Maps in that we wish to create a predictive set of data that will help us determine when to drive on certain roads, but our data is more specific than GoogleâĂŹs broad strokes studies.

*The full version of the author's guide is available as `acmart.pdf` document

BOULDER, 18, March 2018, Boulder, Colorado USA

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

2.2 Data Set

https://www.kaggle.com/jboysen/us-traffic-2015/feed

The dataset is a comprehensive view of various factors which affect the traffic in a particular area at a particular time. Along with the date and times the data was taken at, the data assess factors such as the direction being traveled in, what kind of road is being driven on, the state code, and the traffic volume at various times. There are over thirty attributes and over seven million data points in this dataset that can be used for creating interesting models about our topic.

2.3 Evaluation Methods

After evaluating our dataset we hope to be able to have developed a few different models that will tell us important statistics about what we decide to model. We will try to find out which month that rural roads may be used the less, possibly dependent on the amount of lanes they have, what time of day, and how many cars are on the other roads. These specific results can be used by local governments to plan construction projects at the right time to impact less people.

2.4 Tools

Our project will utilize Python as its primary programming language to perform the data mining processes. We will use Anaconda and ipython notebooks as an easy way to create scripts and manage them. Python is extremely useful in its collection of open source libraries for data mining. This includes pandas, matplotlib, and numpy as they provide programming to manipulate the data and display it visually. Github will be an important asset towards achieving our goal because it serves as our version control as well as our software for project tracking. It will be a place to know what portions of the project needs to be done and what has been done. Other tools we intend to employ would be any other external source we use to help guide and tutor us in the data mining process.

2.5 Milestones

We hope to have the data cleaned and pruned of unnecessary attributes and ready for our analysis to begin by the end of spring break. From there we plan on sorting the data in time for the progress report. Normalization should be done the week after that, along with the creation of our initial models with some preliminary results. In the final weeks we will perfect our models and find our interesting results, as well as finding a way to visualize our results in a meaningful way.

2.6 Peer Review Summary

During the peer review session, there was a general consensus that our projects problem statement, method, and knowledge application were concise and detailed. An improvement to our presentation was that the tools we plan to use for data mining could be more specific as to what purpose they would serve in achieving our goal of the project. Another feedback we received would be to provide a list of the attributes we plan to use for our project and the ones we clean out. Related feedback from other presentations are to find detailed previous work done on our subject and how we can use their results to help crosscheck and analyze our results alongside theirs. A last suggestion or thought is to possibly bring in a new dataset such as weather or RTD or something else that could potentially help explain why we are seeing the traffic trends that we see.

2.7 References

https://www.kaggle.com/jboysen/us-traffic-2015/feed
https://www.nature.com/articles/srep37300
https://www.bactrack.com/blogs/expert-center/35042821-the-most-dangerous-times-on-the-road
https://www.zmescience.com/research/technology/google-maps-traffic-05443/
http://kalw.org/post/driving-apps-waze-are-creating-new-traffic-problems

3 CONCLUSIONS

At this time our group has a good layout of what we want to complete this semester and how we will do it. We plan to find interesting results from our data mining project and be able to create a real report that a city could potentially use to aid in construction projects or city planning around traffic networks. Depending what we find in our analysis we might be able to extend our future plans with our results.