Understanding the Factors for Accident Severity Through Predictive Modeling

Nathan Rugge, Robert King, Anna Lan, and Kareem Naguib
{nrugge3,robert.f1.king,alan3,knaguib3}@gatech.edu

1 TEAM 233 PROJECT PROPOSAL

Today, first responders are called to the scene of an accident based on the description provided by the 911 caller. After classifying the incident, first responders are dispatched and the scene is managed to restore the roadways as safely and quickly as possible. The operator faces the challenges of inaccurate reports and overload from multiple callers – determining the traffic severity of the accident typically requires additional verification[8]*2. In addition to the problem the operator faces, transportation agencies have different criteria for determining the accident severity and its categorization - there isn't a common classification scheme[6]*2 Traffic severity is a key component used across agencies for determining the classification of the incident [6] - this impacts the response, traffic control planning, resource optimization, and operational improvements[9].

Our project intends to provide a tool that helps people understand and explore the factors of accident severity on traffic*1. The tool will provide interactive visualization and predict severity based on immediately available information including time of day, weather conditions, and similar accidents in the area, as well as using features unique to the location itself (e.g., whether there is an amenity, railway, roundabout or even speed bump nearby)*3. The impact of our study should provide benefits to first responders, transportation agencies, roadway designers and GPS developers and users alike, to mention a few*4.

First responders could potentially improve response decisions in presence of missing information or overload of information from multiple callers. Success can be measured through response time and site cleanup. Roadway designers can identify accident-prone locations (as well as areas prone to not having accidents, where this information could be used to remediate the aforementioned accident-prone locations). Success for a roadway designer would be measured through relative accident frequencies post. For GPS developers and users, accuracy of time estimates can be measured, with improvement expected by leveraging our algorithm*5.

Our project risks primarily consist of inability to discover new findings that would bring benefits to our stakeholders*6. The project team has identified free dataset(s) and will be able to conduct the analysis with minimal or no cost via leveraging team members computing resources and open source software packages*7.

The project timeline will run over six weeks between 28 February 2020 and 17 April 2020 with milestones expected to be completed as follows*8. It is expected that there may be instances of rework or shifts in the analysis trajectory based on findings during the analysis. Success at each of the milestones will be measured by completion of each interim deliverable, with results shared with all team members*9. It is expected that responsibility for all deliverables will be shared equally by team members, but ownership of specific items may be assigned as the project progresses.

- `28 February`: Project Proposal submitted.
- `13 March`: Exploratory Data Analysis completed and results reviewed by team members. Analysis methodology chosen and assigned.
- `27 March`: Analysis completed and results reviewed by team members. Final Report sections assigned.
- `17 April`: Final Report completed and submitted.

All team members have contributed a similar amount of effort into developing the project.

2 LITERATURE SURVEY

Moghaddam et al [11] trained artificial neural networks to understand drivers of severities of accidents occurring on the highways of Tehran, Iran. Their study had a high degree of accuracy but noted many driver-specific characteristics and actions that led to higher severities. These data points would not be available for first responders, a focus of our study.

*Each number corresponds to the Heilmeier Questions.

Beshah and Hill [7] studied accident severities in Ethiopia, using a variety of classification models including decision trees, naive Bayes, and KNN. Their findings dialed in on road design, but seemed very limited in terms of data they were able to source. It likely will not be useful for our project, which should be able to deliver a more meaningful output given the additional data available to us.

Garrido et al [2] looked at factors impacting the severity of injury in accidents in Portugal. Their study is useful for our project and notes some predictors and a methodology that we can start with in our analysis. The main shortcoming of their study was a limited dataset of just 4,528 records, which could easily lead to overfitting issues and unjustified extrapolation. Our dataset includes millions of records, providing a significantly more credible base.

Schnyder et al [13] dug into drivers of PTSD symptoms following severe injuries sustained in accidents. It is not useful for our project, as it did not include any information on outside factors that resulted in the accident that caused these injuries. While accident-related injuries are related to our study, our focus is on using readily available information to predict traffic delays from accidents.

Zong et al [15] used an automobile accident dataset to determine which model worked better for predicting severities, the Bayesian network or regression models, ultimately finding the former to be superior. This finding is useful for our study and can help guide our selection of models to use. Our study will improve upon this by using a far bigger dataset, as well as identifying drivers of accident severity, as opposed to only looking at which modeling approach produces better results.

Ma et al [1] took a novel approach to modeling accident frequencies, which led to recommendations for safety treatment and design of highways. Knowledge of the types of road design associated with higher accident rates will benefit our analysis, and we intend to take this a step further to understand the impact these have on severities, as well as uncovering additional drivers.

Al-Ghamdi [4] took a binary view to studying accident severities, limiting his analysis to comparing fatal versus non-fatal accidents using logistic regression. His approach is not going to be useful for our project, as we intend to consider more than two possible outcomes. Further, he had access to just 9 independent variables to consider, limiting his scope considerably compared to what we should be able to achieve.

Wang et al [14] combined accident frequency and severity models to identify sections of roadway that were prone to accidents, ranking their results by monetary cost of accidents by location. This study will be beneficial to our project, as the general approach and variables considered should also be valuable and predictive in our analysis. Its scope was fairly limited, looking at 12,254 accidents in the London vicinity. Our study will be much larger in scale.

Alkheder et al [5] used an artificial neural network to predict accident severities in Abu Dhabi using 4 severity groupings: Minor, Moderate, Severe, and Death. While their approach and variables considered could be leverage for our project, their model accuracy fell off substantially as accident severity increased. The additional data in available for our study should lead to much higher levels of accuracy.

Ratanavaraha and Suangka [10] modeled severities of accidents occurring on highways in Thailand. Their main finding was that an increase in speed led to an increase in accident severity. Speed should be a valuable predictor in our model as well, but we expect to find more than just one variable associated with accident severities.

Sameen and Pradhan [12]improved on previous artificial neural network models for predicting accident severity by leveraging recurrent neural networks, which are more effective when your data is sequential. This result could be useful for our project, insofar as providing a model framework to explore. The focus of this paper was more about finding the right model for predicting severities as opposed to the predictions themselves. We expect our model will provide more valuable insight and potential social benefit with a focus on identifying drivers that could be mitigated.

Abbas [3] studied causes of accidents occurring in Egypt, with a goal of recommending improvements to lower their overall frequency. Abbas specifically tested whether his data could be combined, or if each roadway needed to be studied separately. He found the former to be the case, which is useful for our project. We will improve on his findings by going beyond accident counts to prediction of severities, giving our results more practical applications.

REFERENCES

[1] 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis Prevention* 40, 3 (2008), 964 – 975. https://doi.org/10.1016/j.aap.2007.11.002

[2] 2014. Prediction of road accident severity using the ordered probit model. *Transportation Research Procedia* 3 (2014). https://www.sciencedirect.com/science/article/pii/S2352146514002701

[3] Khaled A Abbas. 2004. Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt. *Accident Analysis Prevention* 36, 2 (2004), 149 – 163. https://doi.org/10.1016/S0001-4575(02)00145-8

[4] Ali S Al-Ghamdi. 2002. Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis Prevention* 34, 6 (2002), 729 – 741. https://doi.org/10.1016/S0001-4575(01)00073-2

[5] Sharaf Alkheder, Madhar Taamneh, and Salah Taamneh. 2017. Severity Prediction of Traffic Accident Using an Artificial Neural Network. *Journal of Forecasting* 36, 1 (2017), 100–108. https://doi.org/10.1002/for.2425 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/for.2425

[6] Kevin Balke, David Fenmo, and Brooke Ullman. 2017. *Incident Management Performance Measures*. Technical Report FHWA-HOP-10-050. Incident Management Performance Measures.

[7] Tibebe Beshah and Shawndra Hill. 2010. Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia. *AAAI Spring Symposium - Technical Report*.

[8] Jodi L. Carson. 2010. *BEST PRACTICES IN TRAFFIC INCIDENT MANAGEMENT*. Technical Report FHWA-HOP-10-050. Texas Transportation Institute.

[9] Richard Margiotta, Rick Dowling, and Jawad Paracha. 2012. *Analysis, Modeling, and Simulation for Traffic Incident Management Applications*. Technical Report FHWA-HOP-12-045. Cambridge Systematics, Inc.

[10] Vatanavongs Ratanavaraha and Sonnarong Suangka. 2014. Impacts of accident severity factors and loss values of crashes on expressways in Thailand. *IATSS Research* 37, 2 (2014), 130 – 136. https://doi.org/10.1016/j.iatssr.2013.07.001

[11] F. Rezaie Moghaddam, Sh. Afandizadeh, and M. and Ziyadi. 2011. Prediction of accident severity using artificial neural networks. *International Journal of Civil Engineering* 9, 1 (2011). arXiv:http://ijce.iust.ac.ir/article-1-544-en.pdf http://ijce.iust.ac.ir/article-1-544-en.html

[12] Maher Ibrahim Sameen and Biswajeet Pradhan. 2017. Severity Prediction of Traffic Accidents with Recurrent Neural Networks. *Applied Sciences* 7, 6 (2017). https://doi.org/10.3390/app7060476

[13] Ulrich Schnyder, Hanspeter Moergeli, Richard Klaghofer, and Claus Buddeberg. 2001. Incidence and Prediction of Posttraumatic Stress Disorder Symptoms in Severely Injured Accident Victims. *American Journal of Psychiatry* 158, 4 (2001), 594–599. https://doi.org/10.1176/appi.ajp.158.4.594 arXiv:https://doi.org/10.1176/appi.ajp.158.4.594 PMID: 11282694.

[14] Chao Wang, Mohammed A. Quddus, and Stephen G. Ison. 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. *Accident Analysis Prevention* 43, 6 (2011), 1979 – 1990. https://doi.org/10.1016/j.aap.2011.05.016

[15] Fang Zong, Hongguo Xu, and Huiyong Zhang. 2013. Prediction for Traffic Accident Severity: Comparing the Bayesian Network and Regression Models. *Hindawi Publishing Corporation* (2013). https://www.hindawi.com/journals/mpe/2013/475194/