Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

TRAFFIC FORECASTING: A HYBRID APPROACH USING SIMULATION AND MACHINE LEARNING

Casey Bowman

Department of Mathematics
University of North Georgia
1201 Bishop Farms Pkwy
Watkinsville, GA, USA

ABSTRACT

The forecasting of traffic conditions is of great concern to many people around the world every day. The prediction of travel times based on historical data, road congestion, weather, and special events helps not only individuals in their daily driving, but also helps companies and government entities in their planning as well. Many approaches have been used for traffic forecasting, from time series analysis to artificial intelligence techniques to simulation. We propose that combining approaches through machine learning can yield better results than each model can produce on its own.

1 INTRODUCTION

Simulation, as well as statistical models such as time series and regression, and machine learning techniques such as neural networks (NN), have been used for many years in the realm of traffic analysis. Statistical and machine learning methods take a deep dive into finding and exploiting relationships in the data. Microscopic traffic simulations model the flow of vehicles and how they interact with each other. These interactions, and the physics that control and influence them, are absent from data-only approaches. We propose that combining traditional statistical methods with simulation using machine learning and artificial neural networks can result in better overall results as both approaches have their strengths and weaknesses. To our knowledge there have not yet been any other attempts to combine simulation techniques with forecasting techniques in the context of traffic forecasting.

2 STATISTICAL AND MACHINE LEARNING MODELS

Various techniques have been applied to the problem of traffic forecasting in the past. (Cools et al. 2009) used ARIMA and SARIMA models to investigate traffic counts, and (Lippi et al. 2013) compared several methods including SARIMA, support vector regression (SVR), and NN. (Tan et al. 2009) combined ARIMA, MA, and exponential smoothing models using a neural network. Many different techniques were implemented and compared by (Peng and Miller 2019), including statistical models such as time series, exponential smoothing, and regression models, as well as machine learning models such as SVR, and various forms of NN such as feedforward, extreme learning machine (ELM), and long short-term memory (LSTM). They focused on both univariate and multivariate scenarios. They conclude that machine learning techniques and regression models tend to perform better than traditional time-series models, and that multivariate approaches perform better than univariate ones.

3 TRAFFIC SIMULATION

We have chosen a microscopic simulation approach, using our own model for vehicle arrivals to the traffic system, and which will use one of several established car following models such as the Intelligent Driver Model (Treiber et al. 2000), and Gipps' model (Gipps 1981) . These models have been incorporated in our simulations, which have been created using ScalaTion (Miller et al. 2010).

Vehicle count data collected at many sensors in the Caltrans Permformamce Measurement System (PeMS) system in California are used to create a model of several stretches of highways around San Diego. Source sensors are identified and their data are used to create a realistic model for vehicle arrivals to the simulation using polynomial regression and a non-homogenous poisson process (Bowman and Miller 2016). Other sensors' data are used to validate the car-following model by measuring them against simulated counts at those same distances along the highway.

4 COMBINING MODELS

Simulation offers an opportunity to include real physical models of vehicles and their interactions on the road to the forecasting. However, simulations can take a long time to run, which is fine for long-term forecasting models, but diminishes the opportunity for simulation to contribute to short-term forecasting systems. Our solution to this problem is to create a simulation with our interarrival model and car-following model, optimize parameters using vehicle count data, and then use the simulation to train an ELM to mimic the output of simulation, resulting in closely matching output with much faster runtimes.

Another method that can be used to combine models is stacking, which brings together multiple learners as inputs to a meta learner. (Wolpert 1992). In this case traditional forecasting techniques and simulation, or an ELM trained to mimic one, are used as the inputs for some type of NN, which is then trained, and used as a stacked learner for the traffic forecasting problem. Vehicle counts are used as training data, where the source sensors act as input, and downstream sensors are the output. One set of training data is used to train the predictive models, and another set is used to create predictions, which, along with the real values, is used as training data for the meta learner.

REFERENCES

Bowman, C. N., and J. A. Miller. 2016. "Modeling Traffic Flow Using Simulation and Big Data Analytics". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 1206–1217. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cools, M., E. Moons, and G. Wets. 2009. "Investigating the Variability in Daily Traffic Counts Through Use of ARIMAX and SARIMAX Models: Assessing the Effect of Holidays on Two Site Locations". *Transportation Research Record* 2136(1):57–66.

Gipps, P. G. 1981. "A Behavioural Car-Following Model for Computer Simulation". *Transportation Research Part B: Methodological* 15(2):105–111.

Lippi, M., M. Bertini, and P. Frasconi. 2013. "Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning". *IEEE Transactions on Intelligent Transportation Systems* 14(2):871–882.

Miller, J. A., J. Han, and M. Hybinette. 2010. "Using Domain Specific Language for Modeling and Simulation: Scalation as a Case Study". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, , and E. Ycesan, 741–752. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Peng, H., and J. A. Miller. 2019. "Multi-step Short Term Traffic Flow Forecasting Using Temporal and Spatial Data". In *International Conference on Big Data*, 110–124. Cham: Springer.

Tan, M.-C., S. C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang. 2009. "An Aggregation Approach to Short-Term Traffic Flow Prediction". *IEEE Transactions on Intelligent Transportation Systems* 10(1):60–69.

Treiber, M., A. Hennecke, and D. Helbing. 2000. "Congested Traffic States in Empirical Observations and Microscopic Simulations". *Physical review E* 62(2):1805–1824.

Wolpert, D. H. 1992. "Stacked Generalization". *Neural networks* 5(2):241–259.