Enhanced short and longer term network performance prediction capabilities through data-driven analytics and simulation:

Short-term Traffic Speed Prediction for Perth Roads Using Machine Learning

EXECUTIVE SUMMARY

Submitted to	Main Roads Western Australia & iMOVE CRC
Prepared by	Chris Bartley, Chao Sun, Mark Reynolds, Wei Liu and Sharon Biermann
Submitted by	Chao Sun (chao.sun@uwa.edu.au), Planning and Transport Research Centre (PATREC)
Steering Committee	Kamal Weeratunga, Steve Atkinson, Graham Jacoby and Chao Sun
Project details	iMOVE ITS Project 1-003, Sub-project 1 (Part 1), in part-fulfilment of Milestone 4
Date	16 January 2020
Version	Final

Big data is driving a rapid digital transformation in the transport industry. The Western Australian Auditor General (2015) pointed out that '*Having consistent, real time information is key to optimising network performance and informing strategic and operational decision-making* (p.20).' However, big data not only brings opportunities but also challenges. Laney (2001) introduced the commonly used three defining dimensions of big data – the 3Vs (volume, variety and velocity). It is not only the *amount* of data, but also the *variation* between datasets and the *speed at which* the data is generated. Others have later extended the concept and added more 'V's but essentially big data is being generated at a *rate* that surpasses human's ability to make sense of it and with levels of *noise* that make interpretation difficult. Therefore, building decision support systems that automate the analytics becomes a necessity.

Data is worthless unless it can be turned into actionable information. For Main Roads Western Australia's (Main Roads) Network Operations Directorate this can be done in two ways: *offline* analysis of historical performance and *online* application to aid real-time operations. The former has been successfully addressed by the Network Performance Reporting System (NetPReS), and the latter is the focus of this report.

This iMOVE sub-project focuses on short-term prediction of average speed for individual road sections in the Perth metropolitan road network up to a horizon of 75 minutes in advance. It aligns well with Network Operations' vision of '*predict in 20 (min), act in 5 (min), change the future'*. Although it will require substantial investment in data infrastructure, good short-term predictions could enable Main Roads to take a proactive approach to network operations, such as stopping gridlocks before they appear and preventing queue spillbacks. It would also enable faster incident detection and recovery.

This investigation used machine learning techniques that *learn from the past* to *predict the future*. The hypothesis was that machine learning could extract hidden value from the Main Roads datasets that would improve prediction performance over naïve and traditional approaches. The results show that our predictive models are robust and perform well against the benchmarks. The highlight is that the accuracy does not decrease dramatically with an increasing prediction time horizon (how far into the future the model predicts), e.g. during the AM and PM peaks, predicting 15 minutes ahead will produce an average percentage error about 9% while for 75 minutes it is just above 10%. The performance gap with benchmarks becomes more pronounced with increasing timespan (Figure 3 & Figure 4).

Long term, we envisage a traffic '*now-casting*' decision support system for network operations. The concept of '*now-casting*' refers to '*the prediction of the present, the very near future and the very recent past*' (Banbura et al. 2012, p196), a term borrowed from meteorology and economics. '*Predicting the present or the very recent past*' is needed because of the delay in data acquisition and cleaning; and the dynamic nature of network operations makes predicting the very near future valuable. Knowing what is likely to happen could help traffic operators making more evidence-based decisions.

The project required numerous rounds of data cleaning, manipulation, modelling and experimentation. A data analysis pipeline was developed to ensure this process was repeatable, reliable and reproducible. It allows data transformations and modelling frameworks to be reused for rapid development when new datasets become available in the future. This pipeline will be handed over to Main Roads.

Given that there is no perfect single data source measuring traffic performance, we recommend further improvement on data quality to address consistency and accuracy issues. It is being partially addressed by another PATREC project on data fusion which is funded by Main Roads.

The research presented in this report delivers on Sub-project 1 (Part 1) of a larger research project comprising two sub-projects:

- Sub-project 1: **Data-driven empirical models for short-term traffic prediction (Part 1)** & non-route-based area optimisation of network productivity (Part 2)
- Sub-project 2: Simulating the traffic impact of AVs and CAVs to Perth's freeways and arterial roads

This research is funded by PATREC and the iMOVE CRC and supported by the Cooperative Research Centres program, an Australian Government initiative. The contribution of the steering committee throughout the project, in guiding, monitoring and review, is gratefully acknowledged.