STAR: **S**mart **T**raffic **A**daptive **R**egulation

Applied Optimization Report
(Sambit Sahoo, Aditya Sinha, Adheesh Trivedi)

March 14, 2025

Contents

1 Introduction

1.1 Motivation

Growing up in a bustling urban city, I never questioned the efficiency of traffic signals. With lanes equally busy throughout the day, fixed-time signals worked seamlessly, ensuring smooth traffic flow. Every cycle felt justified, balancing the needs of all directions without unnecessary delays. The system, though rigid, rarely felt inconvenient because congestion was evenly distributed.

That perspective changed when I visited a small city with uneven traffic distribution. Some lanes were nearly empty, while others had long queues, yet the traffic lights followed the same fixed schedule. I found myself stuck at a red light for over 240 seconds, waiting for an empty road to "clear." It was frustrating and inefficient—clearly, the system wasn't designed for such imbalances. That moment sparked my curiosity: how could traffic signals be optimized to adapt dynamically, minimizing unnecessary wait times while keeping intersections efficient?

1.2 Problem statement

This project aims to optimize signal change timings by leveraging predictive modeling to dynamically adjust traffic light durations based on real-time and historical traffic patterns. By accurately forecasting traffic flow and signal transitions, the proposed system aims to minimize waiting times across different lanes, leading to smoother traffic movement and reduced delays at intersections.

2 Exploration

We explored various works related to our topic of interest. We found a paper that uses reinforcement learning techniques to address traffic light optimization [5]. While we initially considered this approach, our limited formal understanding of reinforcement learning led us to set it aside in favor of alternative methodologies. Consequently, we continued our search for other relevant studies.

2.1 Paper under consideration

Out of all the papers we reviewed, we selected this paper titled *"A Machine Learning Method for Predicting Traffic Signal Timing from Probe Vehicle Data"*[4] because of its practical applicability. It leverages deep learning and SUMO for traffic simulation, which aligns with our interest in exploring simulation techniques. The combination of these approaches makes the study both innovative and implementable in real-world scenarios.

In the paper, they mentioned use of state-of-art machine learning techniques like tree-based ensemble learning models such as random forest (RF), gradient boost machine

(GBM), and extreme gradient boost (XGBoost), for volume estimation by learning from combined dataset of commercial probe data and other infrastructure attributes such as number of lanes, speed limit, and weather. According to their results, the methods were able to provide hourly volume estimates 24 hours a day, 7 days a week, and 365 days a year with around 18% mean absolute error to true volume and about 5% of error with respect to roadway capacity [2].

They also discuss about the two new adaptive traffic signal control algorithms (iterative & optimized signal control algorithm) are proposed based on data from probe vehicles to realize the coordinated signal control of arterial roads on a microscopic simulation platform. Results show that the average travel time is reduced by 32% under the iterative signal control algorithm and by 23% under the optimized signal control algorithm, and the average delay times are reduced by 36% and 35%, respectively. The average number of stops under the iterative signal control algorithm is reduced by 43%, and under the optimized signal control algorithm, by 67% [3].

3 Literature Review

3.1 Methodology followed

The study follows a structured machine learning pipeline for feature extraction, model training, and evaluation:

3.1.1 Data Processing:

- The researchers extracted **Floating Car Data (FCD)** from SUMO simulations and filtered vehicle trajectories passing through intersections.
- A bounding box of 500 feet around each intersection was used to capture relevant traffic movements.
- Only vehicles that came to a stop before passing through an intersection were considered for estimating red light durations.

3.1.2 Feature Engineering:

- The cycle length estimation process involved detecting acceleration start times after a stop and analyzing their periodicity.
- Fourier Transform techniques were used to identify dominant frequency components in vehicle acceleration start times, which served as input features for the XGBoost model.
- For red time estimation, vehicle stop durations were binned based on intersection location, direction of travel, and time of day. The data was further augmented using empirical quantiles to improve model generalization.

3.1.3 Model Training and Optimization:

- The XGBoost model was trained on Fourier-transformed acceleration time features to estimate cycle lengths.
- A dense neural network (DNN) with 11 hidden layers was trained to predict red light durations using 100-dimensional quantile vectors.
- Bayesian optimization techniques were employed to fine-tune the hyperparameters of both models, ensuring minimal error.
- The training dataset was split into an 80% training set and 20% validation set, with cross-validation applied to prevent overfitting.

3.2 Dataset

In the papers [5, 1], a simulated road network is used to generate high-fidelity **probe vehicle data**. The simulation records second-by-second vehicle trajectories and speeds, which are then processed to extract features such as acceleration start times, stop durations, and deceleration profiles. A bounding box is typically applied around each intersection to ensure that only relevant trips—those that interact directly with the traffic signal—are included in the analysis.

3.2.1 Synthesis of dataset

The significant progress was made in the data preparation and analysis phase of the traffic light optimization project. The focus was on extracting vehicle counts from the Syntrac dataset[1], organizing the data, and preparing it for optimization.

To count the number of vehicles faster R-CNN was used which detected and counted vehicles in each lane for every image. The counts were stored in a structured CSV file, with each row representing the number of vehicles (a,b,c,d) in the four lanes of an intersection. Since the Syntrac dataset presents one problem in front of us of having data for a single lane only in different different scenarios. To tackle this problem and to simulate real-world traffic scenarios, the dataset was randomly sampled into groups of four images(considering the real life 4 lane intersection), ensuring unbiased representation of traffic patterns. These sampled groups will serve as input for the optimization model. This approach allows for a diverse range of traffic conditions to be considered in the optimization model, enhancing its robustness and applicability to real-world situations.

The sampled data was then analyzed to understand the distribution of vehicles across lanes and to identify any patterns or outliers. This analysis is crucial for designing the optimization model and validating its effectiveness. By examining the variability in traffic patterns, we can ensure that the model is capable of handling different traffic densities and configurations.

The sampled data was then analyzed to understand the distribution of vehicles across lanes and to identify any patterns or outliers. This analysis is crucial for designing the

optimization model and validating its effectiveness. By examining the variability in traffic patterns, we can ensure that the model is capable of handling different traffic densities and configurations.

Throughout this process, several challenges were encountered and addressed. Ensuring accurate vehicle counts from the Syntrac dataset required a combination of object detection techniques and manual verification to improve accuracy. Handling variability in traffic patterns across sampled images was managed through random sampling, ensuring a diverse and representative dataset.

References

[1] Tiejin Chen, Prithvi Shirke, Bharatesh Chakravarthi, Arpitsinh Vaghela, Longchao Da, Duo Lu, Yezhou Yang, and Hua Wei. Syntrac: A synthetic dataset for traffic signal control from traffic monitoring cameras. *arXiv preprint arXiv:2408.09588*, 2024.

[2] Yi Hou, Stanley E Young, Anuj Dimri, and Nicholas Cohn. Network scale ubiquitous volume estimation using tree-based ensemble learning methods. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2018.

[3] Fushi Lian, Bokui Chen, Kai Zhang, Lixin Miao, Jinchao Wu, and Shichao Luan. Adaptive traffic signal control algorithms based on probe vehicle data. *Journal of Intelligent Transportation Systems*, 25(1):41–57, 2021.

[4] Juliette Ugirumurera, Joseph Severino, Erik A Bensen, Qichao Wang, and Jane Macfarlane. A machine learning method for predicting traffic signal timing from probe vehicle data. *arXiv preprint arXiv:2308.02370*, 2023.

[5] Marco Wiering, Jelle Van Veenen, Jilles Vreeken, and Arne Koopman. Intelligent traffic light control. *Institute of Information and Computing Sciences. Utrecht University*, 2004.