Integrating transit signal priority with adaptive signal control in a connected vehicle environment

Phase 1 Final Report

Mobility 21 University Transportation Center
Carnegie Mellon University

Stephen F. Smith (PI),
Isaac Isukapati, Eli Bronstein, and Conor Igoe

June 30, 2018

DISCLAIMER

1. Overview

This project has taken significant steps toward the development and demonstration of a more effective approach to transit signal priority (TSP) than currently available systems, through integration of real-time adaptive signal control technology with DSRC-based detection of buses and communication of status information. Our approach has been to extend the SURTRAC adaptive signal control system (in particular, its core intersection scheduling procedure) to achieve TSP objectives while at the same time minimizing the disruptive effects to overall traffic flow efficiency. Recent work with SURTRAC has produced an intersection scheduling procedure that inputs weights reflective of the relative priority of different types of vehicles and pedestrians, and uses these weights to generate signal timing plans that minimize the cumulative weighted delay of currently perceived incoming traffic flows. Taking this procedure as a starting point, the Phase 1 project has investigated mechanisms for generating and incorporating knowledge of expected bus stop dwell times into SURTRAC's aggregate representation of traffic inflows, to more accurately reflect bus arrival times at the intersection (as well as the arrival times of passenger vehicles that are likely to be blocked during dwell time at the bus stop).

One principal research accomplishment has been the development of a statistical method for generating bus stop dwell time distributions that enable highly predictable real-time predictions of dwell times. Specifically, a hierarchical Bayesian model was defined to provide a method for predicting dwell times with small error thresholds with low quantities of data samples. Using two years of bus stop dwell time data provided by the Port Authority of Allegheny County for two major Port Authority bus routes that run through the SURTRAC controlled corridors in the Pittsburgh East End as test data, the approach has been shown to produce predictions within +/– 5 seconds as high as 80% of the time across several intersections, and to significantly outperform a classical linear regression approach.

A second result has been demonstration of the viability of installing a DSRC On Board Unit (OBU). Working together with personnel at the Port Authority's East Liberty maintenance facility, an initial Port Authority bus was outfitted with a DSRC OBU, and successfully tracked through the SURTRAC connected vehicle test bed through receipt of Basic Safety Message (BSM) broadcast from the bus's OBU. The Port Authority has subsequently agreed to acquire and outfit ~50 buses that move through these same corridors, and in Phase 2 of the project the goal is to demonstrate in the field, improved traffic flow efficiency through integration of both Phase 1 technology results. As a first step in Phase 2, real-time DSRC BSM location information will be combined with the use of generated dwell time distribution models to more accurately predict bus arrival times at the intersection. As a second step, the DSRC OBU will be integrated with the bus's onboard computer, and the additional benefit of using other real-time bus status information, including such factors as bus schedule status (ahead or behind), bus occupancy, and bus door status (open or closed), will be analyzed.

The remainder of this report is organized as follows. In Section 2, the hierarchical Bayesian model approach developed for generating predictable bus dwell time models for specific bus stops is summarized, and experimental results obtained with historical data provided by the Port Authority are summarized. In Section 3, an initial experiment demonstrating the ability instrument a Port Authority bus with a DSRC OBU and track its progress through the SURTRAC connected vehicle test bed is described. In Section 4, major accomplishments and conclusions of the Phase 1 project are summarized and future work is described.

2. Modeling Bus Stop Dwell Times

The presence of transit vehicles stopping on urban streets often restricts or blocks other traffic on the road depending on stop locations, resulting in increased overall wait times and delays throughout the system. As a result, unlike other vehicles, transit vehicle trajectories are rarely well integrated into conventional signal coordination plans. In principle, the trajectory of a transit vehicle can be defined as the sum of stop-to-stop link travel times and the time the transit vehicle dwells at stops. It is widely accepted that the dwell times at stops are a major source of the variability in the stop-to-stop travel times. Therefore, the ability to accurately predict dwell times has major value in predicting the stop-to-stop travel time distributions.

One main goal of the first phase of the project has been to create defensible dwell time predictions using small data sample sizes, specifically demonstrating that Bayesian inference can be employed to reliably estimate stop dwell times with just a few data samples. The efficacy of the Bayesian model was evaluated on data from October 2012 and the results are benchmarked against those obtained from a linear regression model, which is trained at each bus stop on September 2012 data. In the following subsections, we present details on the Bayesian model and summarize the experimental results.

2.1 Model Overview

Constructing a predictive bus dwell time distribution model involves three sub-tasks: 1) choosing the likelihood function for posterior updates; 2) choosing principal covariates that influence dwell time distributions; and 3) formalizing a dwell time model using information from the previous sub-tasks.

2.1.1 Likelihood Function For Posterior Updates

For this task, we used historical data for choosing a likelihood function. Specifically, we used the Port Authority of Allegheny County's (PAAC) Advanced Vehicle Location (AVL) weekday dataset. The data is chronologically ordered, and empirical CDFs based on every fifteen minutes of data are created. Dwell times in the APCC dataset are rounded to the nearest second. To address this, two different continuous empirical CDFs are generated using Gaussian, and Gamma KDE techniques. Next, using the same temporally sequential data six analytic distributions (Non-central F, Burr, Weibull, Beta, Log-normal, and Fisk or Log-logistic) are generated. Max-deviation scores are computed between each analytic distribution fit and each of the two empirical distributions. Based on MDT scores, we chose the Log-logistic (Fisk) distribution as the likelihood for the posterior updates.

2.1.2 Covariates For Dwell Times

In order to develop a dwell time model with covariates, several relationships were explored between covariate data and dwell time, such as the number of onboarding passengers (x_{on}), number of alighting passengers (x_{off}), and load of the bus (x_{load}). A clear positive correlation was found between first two covariates and dwell time, which were chosen as covariates in developing the predictive dwell time distribution model. A scatter plot demonstrating the relationship between the number of onboarding passengers and the dwell time is presented in Figure 1, Figure 2 demonstrates not only that more onboarding passengers corresponds to longer dwell times, but also that the variance of the dwell time increases as more passengers board.

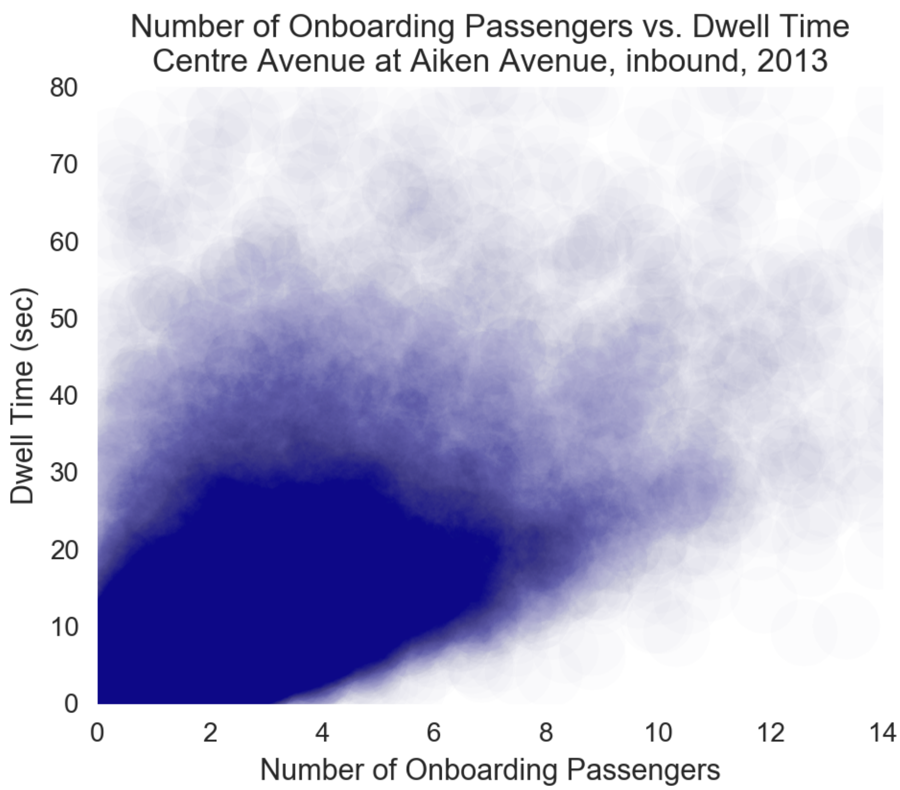

Fig. 1: Scatterplot of # onboardings vs. dwell times

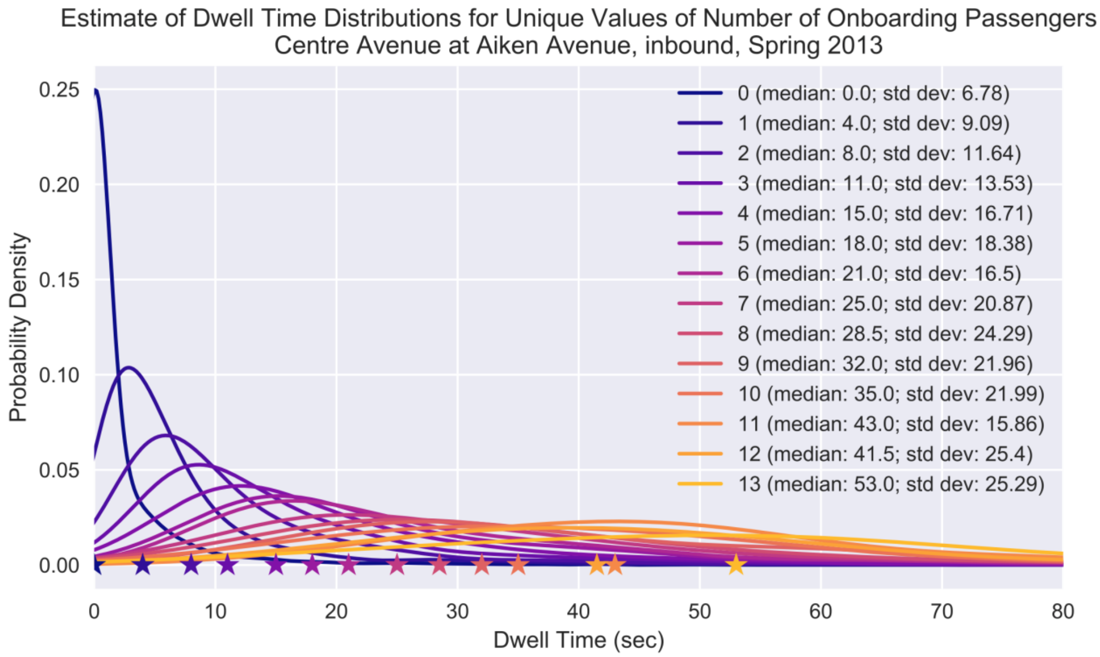

Fig. 2: Conditional dwell time distributions for several numbers of onboarding passengers. Note that the variance is larger when more passengers board.

2.1.3 Dwell Time Model With Covariates

The following describes a Bayesian parametric model for bus dwell times using two covariates $\mathbf{x_{on}}$ and $\mathbf{x_{off}}$. Based on the analysis presented in the subsection on choosing the likelihood function, bus dwell time is modeled as a random variable **X** following a Log-Logistic (Fisk) distribution. Equivalently, bus dwell times **X** are distributed following the exponential of the Logistic distribution. Covariate parameters are introduced by parameterizing the **s** parameter, and the median of the Log-Logistic distribution. The exponential relationship between the Logistic and Log-Logistic distributions is used in this formulation. This parameterization is described below:

$$X = \exp(Y)$$

$$\text{Where } Y \sim \text{Logistic}(\mu, s)$$

$$\mu = \ln(\alpha) = \ln(\beta_\alpha^T x + \beta_0)$$

$$s = {}^1/_\tau = {}^1/_{(\beta_\tau^T x)}$$

$$\beta_\alpha = \begin{bmatrix}\beta_\alpha^{on} & \beta_\alpha^{off}\end{bmatrix}^T$$

$$\beta_\tau = \begin{bmatrix}\beta_\tau^{on} & \beta_\tau^{off}\end{bmatrix}^T$$

$$x = [x_{on} \quad x_{off}]^T$$

At any given time, the belief of the two parameters **μ** and **s** describe current belief of bus dwell time distribution. In a real-time system with access to dwell time observations, belief of the parameter distributions is continuously updated in the light of new data. Bayes' Theorem offers a natural way to achieve such an update scheme. As only one observed dwell time **d** is considered during any Bayesian update, the likelihood function is given by

$$L(\mu, s | \ln(d)) = f(\ln(d), \mu, s)$$

Where f is the probability density function of a Logistic distribution.

Before obtaining any posterior distributions to use as priors, we bootstrap the model using a Normal prior for each of the 4 covariate parameters: β_α^{on}, β_α^{off}, β_τ^{on}, β_τ^{off} and offset parameter β_0. Once a set of posterior distributions is obtained, the most recent posterior distributions are used as priors in the next Bayesian update. The Metropolis Hastings algorithm is employed to obtain MCMC samples of the posterior distributions for four covariate parameters and the offset parameter.

To make a dwell time prediction for an approaching bus, we observe values for covariates x_{on}, and x_{off}, and use posterior distributions of each β to determine the posterior predictive distribution of **X**.

This process is repeated in the light of new data, using the most recent posterior distributions of each β as priors in the next Bayesian update. The means of several model parameters are shown in Figures 3, where a real-time prediction scenario is simulated on historical data in a rolling fashion.

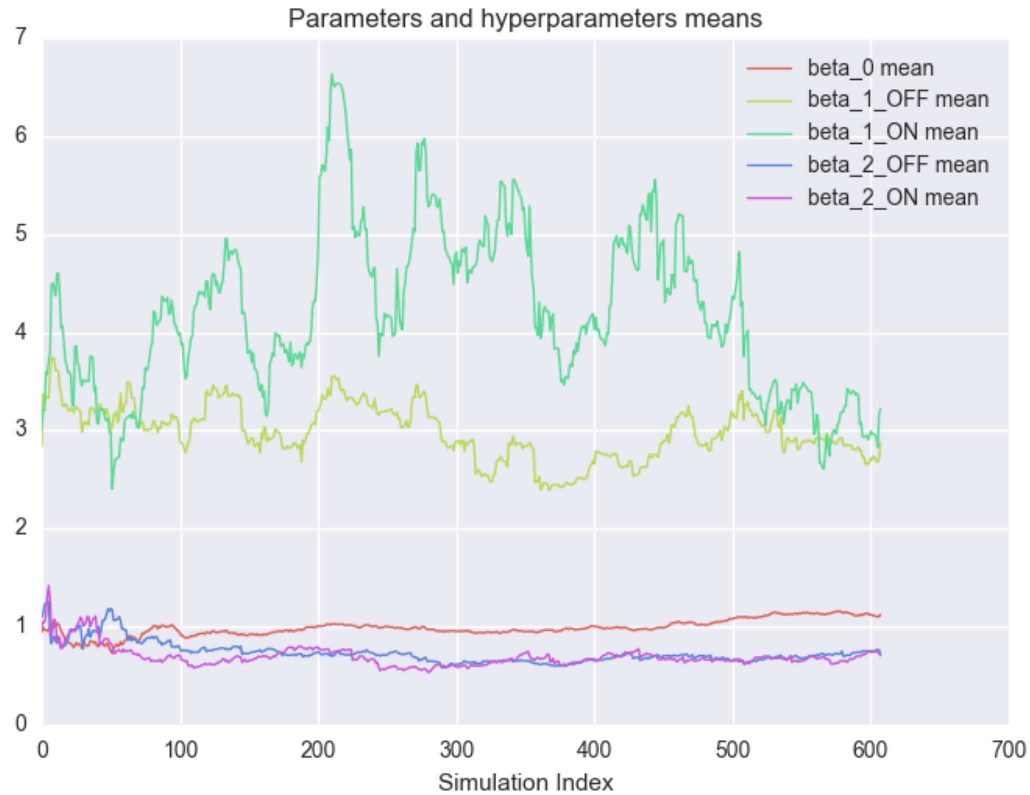

Fig. 3: Means of model parameters throughout simulation. beta_1 corresponds to beta_alpha, beta_2 corresponds to beta_tau

2.2 Model Testing

We tested the efficacy of the proposed dwell time prediction model on bus dwell time data provided by the Port Authority of Allegheny County in Pittsburgh, Pennsylvania for the period from September 2012 to August 2014. While the dataset spans over two years, data from October 2012 is used to test the Bayesian model. We compared the results of the Bayesian model to those of a linear regression model for benchmarking purposes. We trained a linear regression model on September 2012 and tested on October 2012, which are good training and test datasets since it is widely accepted that seasonal trends in bus dwell time distributions are statistically similar. Therefore, the linear regression model is not really put to the test. In principle, regression equations for September 2012 & October 2012 should look very similar, suggesting that predictions on the test dataset should be reasonably good. However, the main objective of this analysis is to evaluate the robustness of the proposed framework. In other words, the goal is to check whether the Bayesian model is able to predict dwell times without any training and how good those predictions are compared to predictions from a well-trained traditional model.

With these objectives in mind, we tested the robustness of the Bayesian framework at twelve different bus stops in the East End region along Centre Avenue corridor in Pittsburgh, PA.

2.2.1 Cumulative density functions of dwell times

Analyzing cumulative density functions (CDFs) of dwell times provides useful insights into the reliability (presence or absence of variance) of these distributions. From the standpoint of stochastic dominance, the distributions with curves furthest to the left have smaller variance in dwell time distributions and hence are more reliable.

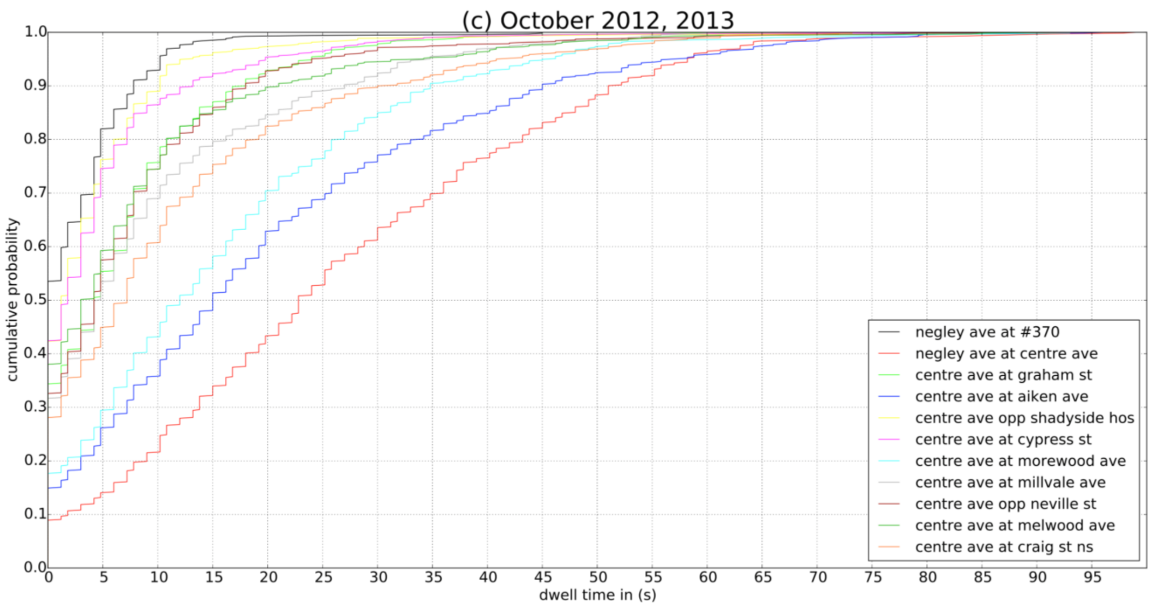

Fig. 4: Cumulative density functions of dwell times

Figure 4 presents dwell time CDFs for test bus stops of interest. It can be seen that dwell time distributions have the largest variance at Negley Ave at Centre Ave (CDF in red), followed by Centre Ave at Aiken Ave (blue), Centre Ave at Morewood Ave (cyan), Centre Ave at Craig St NS (peach), and Centre Ave at Millvale (light grey). This information is useful because predicting dwell time distributions at these intersections is particularly hard due to their highly stochastic nature.

2.2.2 Model Performance

As mentioned earlier, the efficacy of the Bayesian model is evaluated on data from October 2012 and the results are benchmarked against those obtained from a linear regression model, which is trained at each bus stop on September 2012 data. The same Bayesian parametric model is applied to each of the bus stops, and we set Normal priors for each of the 4 covariate parameters and the offset parameter β_0. Covariate parameters are updated on an ex post facto basis, and dwell time predictions are made starting from the very first new data point onward.

We use the ability to predict dwell times within a small error threshold as a performance metric to evaluate the models. The rationale for choosing small error bounds is to account for the fact that these dwell time values are used by planning algorithms in real-time systems, so larger errors will generate schedules that are far from optimal. For this reason, the fraction of predictions within error bounds of [-5, 5] seconds is used as a performance metric. Effectively, this fraction represents the area under the error distribution density function within these tolerance bounds. This is a more informative metric in the context of traffic signal scheduling due to the importance of maximizing the proportion of very close predictions.

Table 1 summarizes performance of these two models. As can be seen, this table contains three sets of performance comparisons: 1) morning peak hour (7:00 - 10:00 AM); 2) evening peak hour (4:00 - 7:00 PM); and 3) the entire test dataset. This table has four columns: first column presents bus stop location information; second column presents fraction of dwell time predictions within the range of -5 and 0 seconds; third and fourth columns contain similar information but for ranges of [0, 5] and [-5, 5] seconds respectively. Lastly, each row contains results for a specific bus stop.

The following inferences can be drawn based on these results: First, the Bayesian predictive model performs at least as good as or better than the linear regression model. This is very encouraging to see as it validates the main philosophy behind the development of this framework, i.e., to develop a predictive probabilistic model for estimating task durations without making use of large training datasets. Second, for the scenarios in which dwell time distributions are highly stochastic (refer Fig 4), the Bayesian prediction model significantly outperforms the linear regression model (refer to results for Negley Ave at Centre Ave, Centre Ave at Aiken Ave, and Centre Ave at Craig St NS). Figure 5 demonstrates this trend for Negley Ave at Centre Ave - the Bayesian model has a much higher proportion of very close predictions than the linear regression error distribution. This again corroborates the hypothesis of quick adaptability of the Bayesian model. Third, in addition to dwell time estimates, the variance or precision parameter of the Bayesian model quantifies the uncertainty of each prediction.

Bus Stop		[-5, 0]		[0, 5]		[-5, 5]	
		L.R.	Fisk	L.R.	Fisk	L.R.	Fisk
Centre Ave at Aiken Ave	AM	0.11	0.22	0.31	0.29	0.42	0.51
	PM	0.11	0.20	0.40	0.44	0.51	0.64
	All	0.10	0.21	0.39	0.41	0.49	0.63
Negley Ave at Centre Ave	AM	0.13	0.18	0.13	0.33	0.27	0.51
	PM	0.08	0.14	0.09	0.33	0.16	0.47
	All	0.10	0.15	0.11	0.32	0.21	0.46
Negley Ave at #370	AM	0.20	0.33	0.59	0.52	0.79	0.84
	PM	0.13	0.21	0.66	0.66	0.79	0.87
	All	0.15	0.29	0.63	0.57	0.78	0.86
Centre Ave Opp Neville St	AM	0.23	0.31	0.45	0.48	0.69	0.79
	PM	0.18	0.26	0.55	0.47	0.73	0.73
	All	0.21	0.29	0.51	0.50	0.72	0.79
Centre Ave at Shadyside Hos	AM	0.25	0.36	0.54	0.45	0.79	0.80
	PM	0.22	0.27	0.48	0.48	0.70	0.75
	All	0.20	0.30	0.52	0.46	0.73	0.76
Centre Ave at Morewood Ave	AM	0.16	0.25	0.38	0.35	0.54	0.61
	PM	0.16	0.37	0.58	0.44	0.73	0.81
	All	0.15	0.28	0.52	0.45	0.67	0.73
Centre Ave at Millvale Ave	AM	0.14	0.25	0.41	0.45	0.54	0.71
	PM	0.18	0.35	0.50	0.44	0.68	0.79
	All	0.13	0.28	0.51	0.50	0.65	0.78
Centre Ave at Melwood Ave	AM	0.21	0.25	0.46	0.51	0.67	0.76
	PM	0.23	0.30	0.54	0.50	0.77	0.79
	All	0.20	0.28	0.54	0.52	0.74	0.80
Centre Ave at Graham St	AM	0.17	0.34	0.57	0.44	0.73	0.78
	PM	0.24	0.34	0.53	0.42	0.76	0.76
	All	0.19	0.31	0.59	0.48	0.78	0.80
Centre Ave at Cypress St	AM	0.19	0.30	0.55	0.47	0.74	0.77
	PM	0.09	0.26	0.53	0.42	0.62	0.68
	All	0.16	0.26	0.54	0.46	0.70	0.73
Centre Ave at Craig St NS	AM	0.07	0.17	0.25	0.32	0.32	0.49
	PM	0.15	0.13	0.26	0.30	0.42	0.43
	All	0.11	0.17	0.26	0.34	0.37	0.51
Centre Ave Opp Shadyside Hos	AM	0.22	0.32	0.60	0.51	0.82	0.82
	PM	0.21	0.28	0.54	0.50	0.76	0.78
	All	0.21	0.29	0.57	0.49	0.79	0.78

TABLE I: Model Performance Comparisons

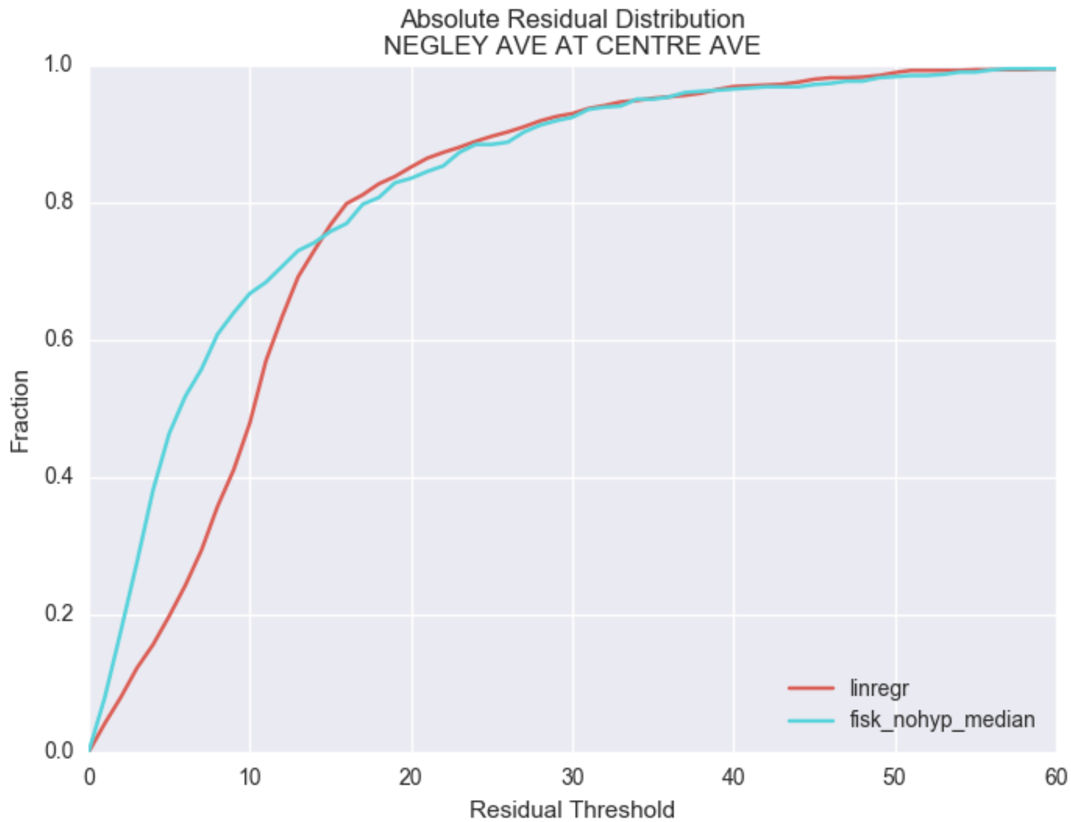

Fig. 5: Fraction of absolute prediction error within a threshold for our framework vs. linear regression. Note that the Bayesian hierarchical model has a higher proportion of small errors.

3. Capturing Real-Time Bus Information

A second thrust of the Phase 1 project focused on demonstrating the feasibility of receiving real-time location information from buses as they approach intersections via DSRC-based vehicle-to-infrastructure (V2I) communication. In cooperation with Port Authority personnel at the East Liberty garage, a Locomate Mini2 OBU was installed over the front window of the bus (see Figures 6 and 7), and connected to an antenna that was attached on the bus exterior right above the route sign (Figure 9). The installation process turned out to be quite straightforward and took under an hour. The viability of this installation approach was subsequently verified.

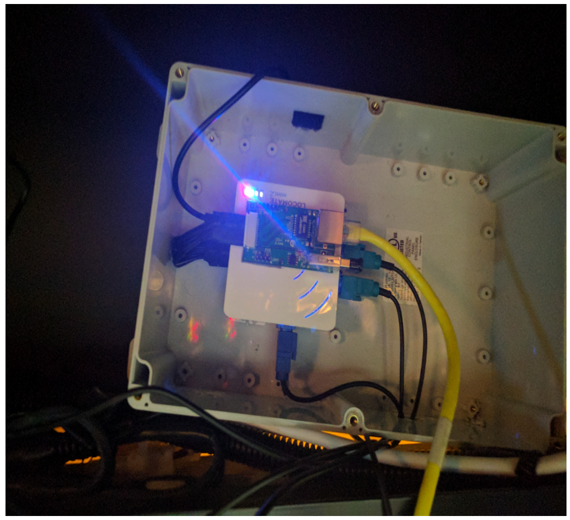

Figure 6: Installed Locomate Mini2 DSRC OBU (in a protective casing)

Figure 7: OBU Install location on shelf above the front window.

Figure 8: OBU antenna location on the bus exterior above the route sign.

To verify the effectiveness of the OBU installation strategy, a bus-tracking experiment was performed using the Map Application of the Connected Vehicle Warehouse Tools, a suite developed under direction of the Federal Highway Administration by the Leidos Corporation for purposes of consolidating efforts of FHWA-approved connected vehicle test beds. Specifically, the on-board OBU was configured to continuously emit Basic Safety Messages (BSMs), and BSM processing code was developed on the Road Side Equipment (RSE) unit side to receive the BSM, and forward message contents to the Leidos visualization tool. The equipped bus was then driven through the Centre Avenue corridor of the SURTRAC connected vehicle test bed.

The sequence of screenshots displayed in Figures 9-11 track the equipped bus's progress moving east along Centre Avenue. As can be seen, the Leidos Map tool shows the real-time traffic signal status in each direction at each intersection (green, red or yellow) as well as the real-time status of equipped vehicles over a specified time window. In the sequence of screenshots shown, the bus is the only equipped vehicle being monitored. In Figure 9, the bus is first picked up while stopped at the bus stop between Cypress and Centre Avenues and then travels to the intersection at Aiken Street. (Within the equipped vehicle trajectory, red discs indicate that the vehicle is stopped or moving < 5 miles per hour; yellow discs indicate the vehicle is traveling between 5 – 20 miles per hour; and green indicates that the vehicle is traveling > 20 per hour.) Figure 10, then shows the bus moving through the intersection of Centre Avenue and Aiken Street, and then finally in Figure 11, the bus is shown further east crossing South Graham Street and heading for the intersection at Negley Avenue. Throughout the tracking experiment the OBU was found to function perfectly.

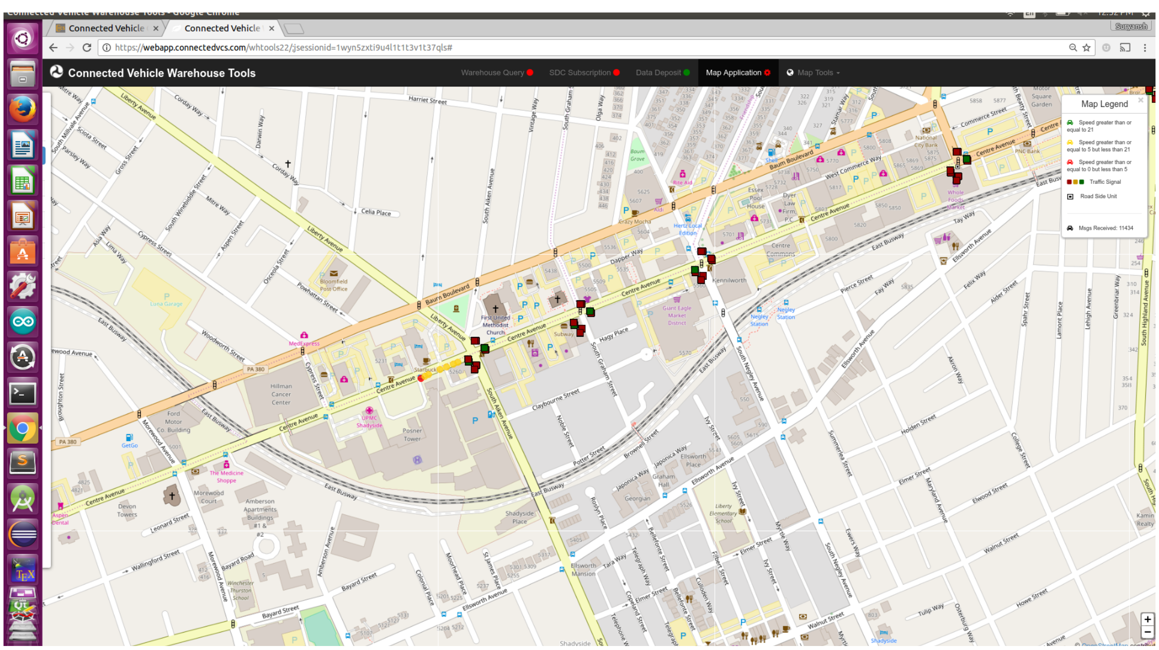

Figure 9: Visualization of Bus Tracking Experiment (1)

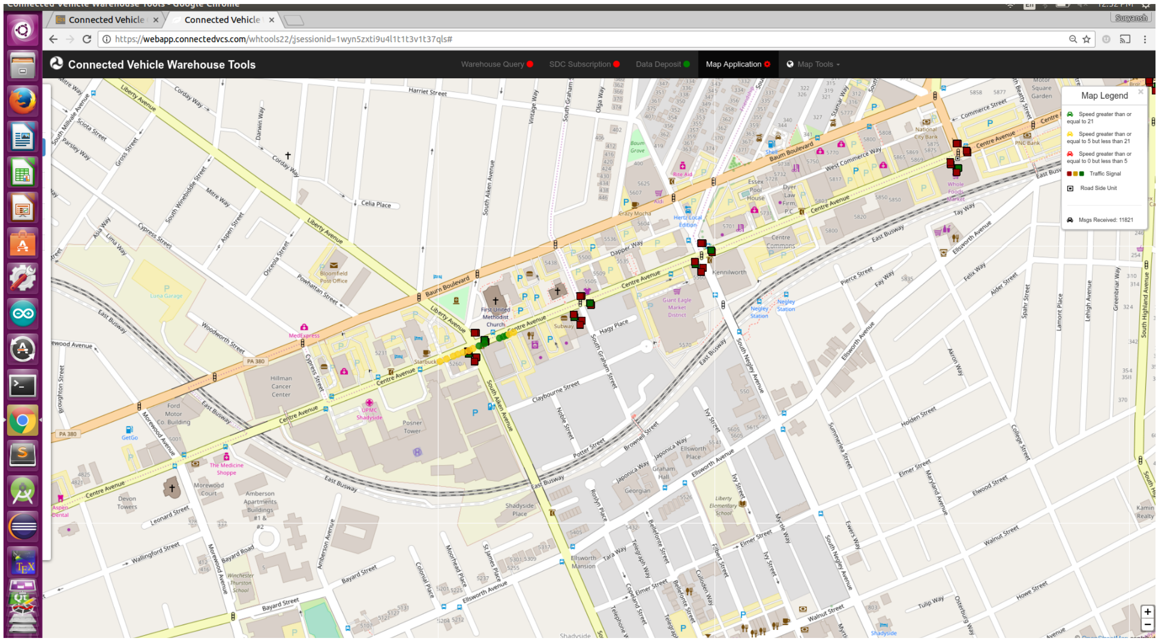

Figure 10: Visualization of Bus Tracking Experiment (2)

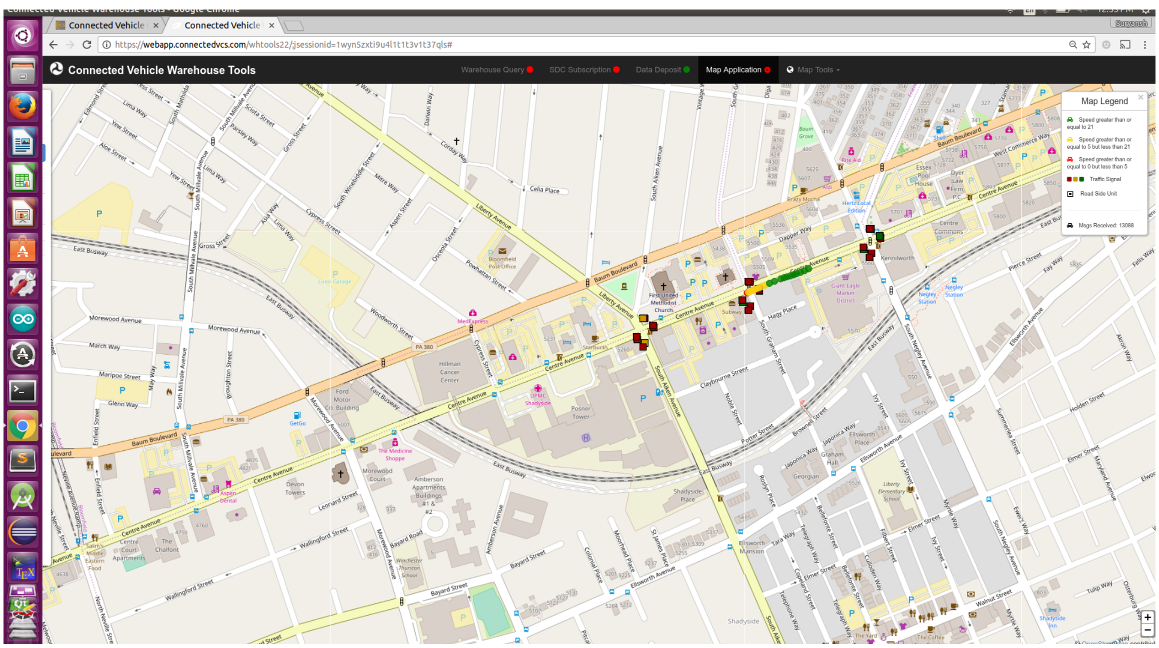

Figure 11: Visualization of Bus Tracking Experiment (3)

4. Conclusions

In this project, we have developed technology components essential to the development of a new approach to transit signal priority – one that exploits real-time adaptive signal control to better optimize bus movements within the context of actual surrounding traffic flows. Where as contemporary transit priority systems give unconditional priority to buses to the detriment of all other traffic on the road, we advocate an approach that moves buses along effectively while balancing the demands of surrounding traffic flows. "Requests" from buses come in via DSRC-based vehicle-to-infrastructure (V2I) communication, and all approaching vehicles are weighted according to mode type. To give active attention to buses they can be assigned a high weight relative to other mode types, and these weights will directly influence real-time optimization of signal timing plans. In the longer term, this basic prioritization scheme can be augmented to incorporate other real-time bus status information such as whether the bus is ahead of or behind schedule. This approach also provides a natural basis for resolving simultaneous competing bus requests for priority, which are typically addressed in an ad-hoc manner at best by contemporary transit priority solutions.

Toward the goal of integrating real-time time signal control with vehicle-to-infrastructure (V2I) communication to achieve more effective transit priority solutions, this project has made significant progress. One major accomplishment has been the development of a predictive Bayesian parametric model for capturing and exploiting bus dwell time distributions. Predicting bus dwell time at intervening bus stops is unquestionably the most challenging aspect of using real-time information on bus location to predict arrival at the intersection. The developed model has a couple of desirable features. First it can be shown to predict with greater accuracy than conventional linear regression modeling methodologies. Second, accurate models can be constructed from a small number of sample points, in contrast to contemporary deep learning approaches. The efficacy of the developed model was tested at twelve different bus stops in the Pittsburgh East end, using two years of bus stop dwell time data provided by the Port Authority for two major bus routes. The results indicated that the Bayesian model performed at least as good and in most instances far better than a traditional linear regression model. Furthermore, predictable results were extractable from the generated model after only 20 or so data samples.

At the vehicle-to-infrastructure (V2I) communication level, this Phase 1 project also demonstrated the ability to receive DSRC Basic Safety Messages (BSMs) from an OnBourd Unit (OBU) installed on a Port Authority bus, and to track its movement through the Pittsburgh SURTRAC connected vehicle test bed.

Looking ahead to the Phase 2 project, future work will first focus on combining these dwell time modeling and BSM processing capabilities to improve SURTRAC's internal prediction of approaching traffic flows and demonstrating the use of this improved prediction to further boost traffic flow efficiency in the Pittsburgh SURTRAC traffic control signal network. To facilitate experimentation in the field, the Port Authority has agreed to outfit buses that run through this network with OBUs. Integration of installed OBUs will also be integrated with the buses onboard computer to provide other real-time bus status information (e.g., how full, which route, doors open and doors close events, behind or ahead of schedule) and a second thread of future research will investigate the added benefit of incorporating this information.

References

[Isukapati 18] Isukapati, I., C. Igoe, E. Bronstein, and S.F. Smith, "Generating Highly Predictive Probabilistic Models Of Task Durations", CMU Robotics Institute Technical Report TR-RI-18-39, June 2018.

Appendix A

[Reprint of: Isaac K. Isukapati, Conor Igoe, Eli Bronstein, and Stephen F Smith, “Generating Highly Predictive Probabilistic Models Of Task Durations”, CMU Robotics Institute Technical Report TR-RI-18-39, June 2018.]

GENERATING HIGHLY PREDICTIVE PROBABILISTIC MODELS OF TASK DURATIONS

Isaac K. Isukapati, Conor Igoe, Eli Bronstein, Stephen F Smith

CMU RI Technical Report #
CMU-RI-TR-18-39

June, 2018

ABSTRACT

In many applications, uncertainty in the durations of tasks complicates the development of plans and schedules. This has given rise to a range of resilient planning and scheduling techniques that in some way rely on probabilistic models of task durations. In this paper we consider the problem of using historical data to develop probabilistic task models for such planning and scheduling techniques. We describe a novel, Bayesian hierarchical approach for constructing task duration distributions from past data, and demonstrate its effectiveness in constructing predictive probabilistic distribution models. Unlike traditional statistical learning techniques, the proposed approach relies on minimal data, is inherently adaptive to time varying task duration distribution, and provides a rich description of confidence for decision making. These ideas are demonstrated using historical data provided by a local transit authority on bus dwell times at urban bus stops. Our results show that the task distributions generated by our approach yield significantly more accurate predictions than those generated by standard regression techniques.

Index Terms -- Task Duration prediction, Hierarchical Bayesian Models, Intelligent Transit Systems, Adaptive Control.

INTRODUCTION

Most practical planning and scheduling problems are complicated by the uncertainty inherent in the tasks that must be performed to achieve stated objectives. An attempt by a robot to pick up an unstable object can have multiple outcomes, only one of which achieves the desired effect and allows a larger plan to move forward. A vehicle traveling from a given pickup location to a given drop-off location may have several different routes to choose among, each of whose duration is variable and dependent on current traffic conditions. In the specific application that motivates the work to be presented in this paper, which is online generation of timing plans for signalized traffic intersections, the ability to reliably predict bus dwell times at near-side bus stops is crucial to optimizing movement of approaching traffic flows.

To cope with uncertainty in task durations and outcomes, a range of techniques for building resilient plans and schedules have emerged. Some techniques have relied on knowledge of uncertainty limits to generate plans that retain temporal flexibility [1 – 4]. Others have exploited probabilistic models of task duration and outcome uncertainty to generate plans or policies that optimize expected behavior [5 – 9]. Still other techniques have used probability distribution information to predict durations within deterministic optimization procedures [10, 11].

The effectiveness of all of these techniques of course depends on the availability of good probabilistic task models. In this paper, we consider this requirement, and focus on the issue of acquiring such models. We propose a Bayesian hierarchical approach for constructing highly predictive probability models from past data. Our approach offers several advantages over traditional statistical learning techniques, including the ability to start making accurate predictions with only minimal past data, to provide robustness in stochastic and noisy systems, and to deliver a confidence in predictions. To demonstrate these advantages, we apply the approach to the above mentioned problem of constructing bus dwell time models along a given roadway. Using historical data provided by a local transit authority, results show that significantly more accurate dwell time distributions can be derived from far less data than is possible with standard linear regression methods.

The remainder of the paper is organized as follows. We first describe our Bayesian hierarchical modeling methodology for constructing task duration models and summarize its advantages. Next, we apply the approach to bus dwell time data to analyze its effectiveness in producing bus dwell time models. We then compare the performance of our Bayesian hierarchical model to a more traditional statistical model. Finally, we summarize the contributions and briefly indicate our future research directions.

Bayesian Hierarchical Framework

Many state-of-the-art dynamic adaptive planning systems employ optimization models to decide how to allocate scarce resources among tasks for optimal performance. These systems typically assume that current unfinished tasks have deterministic completion times rather than taking into account task duration uncertainty. Then, to account for dynamic behavior, optimization models are re-run upon the discovery of new information to generate updated optimal plans. These adaptive strategies tend to be reactive rather than proactive. In principle, quantifying uncertainties in task durations will enable anticipatory or proactive strategies that offer more resilient (and more predictable) plans and schedules.

The goal of this research is to utilize the availability of real-time (or near real-time) covariate task duration data to produce more accurate task models for online planning and scheduling. In this context, there are several challenges. First, the environment can be highly stochastic and change over time, making prediction difficult due to the large variance and dynamic nature of the system. Second, there is often noise and outliers in the data, necessitating a robust approach that is not prone to overfitting. Third, the available training datasets may be small, making models with lots of parameters impractical. Fourth, a confidence in the prediction might be necessary, particularly for control decisions that must gauge the uncertainty of the model. Fifth, the implementation of the model in real-time systems should be computationally efficient. Finally, being able to interpret the model and understand the structure of interactions of the variables might be an important requirement. In the following sub-sections, we introduce a Bayesian hierarchical framework that meets these requirements.

Key Concepts of the Framework

Central to the framework is the concept of a *rolling Bayesian update scheme*. Instead of learning a model from a training dataset, or using historical data from multiple qualitatively similar time intervals, we make predictions using a small set of continually updated model parameter distributions. A fundamental component of the proposed framework, then, involves the use of an appropriate analytical statistical model that is determined offline and subsequently refined online. Such a scheme has several advantages over feature-engineered solutions that rely on subsets of historical data at any given time. In many real-world contexts, task duration constitutes a highly stochastic non-stationary process. Consequently, finding informative historical data for any point in time is a difficult and noise-prone endeavor, yielding little valuable signal for the comparatively complex system design. In contrast, high correlation between task duration model parameters exists between short intervals. As a result, there is significant value in maintaining real-time beliefs of a predictive model and updating continually in the light of new data. This results in a lightweight framework that naturally adapts to underlying non-stationary stochastic process, quickly improves with more observations, and easily generalizes to various task duration prediction scenarios.

A second key concept of the framework is in the *hierarchical* nature of resulting models. While the following sections describe the process of creating a task duration model based on one set of observed covariates and completed durations, the hierarchical aspect of the framework allows multiple predictive models to feed into each other. For example, observed covariates of a lower model could be used to predict another set of variables, which are in turn used higher up the hierarchy as covariates to predict the task duration. Multiple layers of statistical models can be connected in this way to create a more complex hierarchy whilst maintaining clear model interpretability. This concept is illustrated in the application section of this paper, where a hierarchy of predictive models is employed to deliver real-time bus dwell time predictions.

Advantages Of The Framework

There are four primary advantages of using this framework. First, this framework offers more robust predictions in highly stochastic and noisy environments, which often have a large variance and noise in both the independent and dependent variables. Second, the Bayesian framework effectively addresses uncertainty by delivering a confidence in the prediction through the posterior predictive distribution, rather than simply supplying a point estimate. This confidence can be useful in control decisions when deciding how to act based on the reliability of an estimate. Third, the framework requires little data, both in the selection and prediction stages. The selection stage involves choosing the likelihood for the prediction variable and prior distributions for the model parameters, both of which can be computed from a small amount of historical data. In the prediction stage, the model can begin making predictions and updating the posterior distribution in a rolling fashion, removing the need for a ``training" dataset once the model has been determined. Fourth, the model is computationally efficient because analytical conjugate posterior distributions are simply described by their parameters, and non-conjugate distributions can be sampled efficiently using Markov Chain Monte Carlo (MCMC) methods, or nested sampling techniques.

The following sub-sections provide details on the individual steps in using the framework.

Selecting The Likelihood Function For Task Duration

The purpose of this first step is to find an analytic distribution that best explains task duration distributions in the empirical data. This is an important step because if such an analytic distribution is found, the posterior task duration distribution can be found in a computationally efficient manner. Unlike the training stages for many complex statistical models, this analysis step, which involves fitting analytical distributions and assessing their statistical similarities, does not require a large amount of data.

Algorithm 1 Choose Task Duration Likelihood Function

$D \leftarrow$ chronologically ordered task duration data
$(t_i, \delta_i) \leftarrow$ time stamp & task duration of record i in D
$(t_l, t_u) \leftarrow$ lower & upper bounds of time interval
$\eta \leftarrow$ length of time window of interest
initialize $(t_l, t_u) \leftarrow (0, \eta)$
for $(t_i, \delta_i) \in D$ **do**
    $L \leftarrow [\,]$
    if $t_l \leq t_i < t_u$ **then**
        append δ_i to L
    else
        compute empirical CDF F from data in L
        fit n CDFs $F' \leftarrow [F_1, ..., F_n]$ to data in L
        $S \leftarrow$ MDT scores for F & each CDF in F'
        write MDT output $[F, F', S]$ to an output file
        update $(t_l, t_u) \leftarrow (t_u, t_u + \eta)$

Algorithm 1 describes the methodology for choosing a task duration likelihood function. The first step is to chronologically order the task duration data. The next step is to develop empirical cumulative density functions (CDFs) F based on temporally sequential sets of observations that fall within the time window of interest. To ensure tight tracking of time-varying parameter distributions, it is prudent to consider intervals of time consistent with decorrelation of the underlying process. In case the task durations δ_i in the data are discretized (due to rounding errors), use Kernel Density Estimation (KDE) techniques to obtain a continuous CDF. Next, use the same temporally sequential sets of observations to fit analytic distributions. A general guidance in that regard is to consider the following six distributions: Non-central F, Burr, Weibull, Beta, Log-normal, and Fisk (Log-logistic). The next step is to statistically analyze similarities between the empirical CDF and each of the analytic distributions using the Maximum Deviation Test (MDT) [12].

As the name suggests, the maximum deviation test is a statistical technique designed to quantify statistical differences between two probability density functions. The methodology employed here generates a test scores that measures statistical similarity between the empirical and each of the analytic distributions. Here the test-score is nothing but the number of percentile values in an analytic distribution (F_i) that are within a user-defined tolerance bounds of the empirical distribution (F). The analytic distribution with highest test score (s_{max}) is statistically most similar to the empirical distribution. Pseudo-code for the methodology is given in Algorithm 2.

Most non-parametric tests, such as the Kalmagorov Smirnov (KS) test [13], use maximum deviation from the mean as a measure to check for dissimilarity. Therefore, these tests fail to recognize dissimilarities in heavy-tailed, or multi-modal distributions. On the other hand, MDT uses the sum of deviations of every percentile of the distribution as a measure for dissimilarity. This property, in addition to the symmetric nature of the test, makes MDT a very powerful test over either the KS Test or the Kullback-Leibler (KL) Divergence test [14]. Therefore, it is appropriate to use MDT for comparing the empirical distribution with each of the analytic distributions.

Note that, should covariate data not be available a priori for prediction, the steps outlined in this sub-section can also be taken to determine an appropriate analytical model for use in estimating covariates.

Algorithm 2 Maximum Deviation Test

$\epsilon_{tol} \leftarrow$ error tolerance threshold
$F \leftarrow$ empirical CDF
$F' \leftarrow [F_1, ..., F_n] \leftarrow$ CDFs of n analytic distributions
initialize test scores $S \leftarrow [0, 0, ..., 0]$
for $F_i \in F'$ **do**
    for p in $[0, 100]$ **do**
        $\epsilon \leftarrow \frac{F^{-1}(p) - F_i^{-1}(p)}{F^{-1}(p)} \times 100$
        if $abs(\epsilon) \leq \epsilon_{tol}$ **then**
            $s_i \leftarrow s_i + 1$
$s_{max} \leftarrow max(S)$
return F_k corresponding to s_{max}

Setting Priors And Generating Predictions

The next step after choosing the likelihood function(s) is to choose a prior distribution for each parameter of the task duration analytical distribution, and any parameters necessary for other models used in the hierarchy for covariate estimation. This is a fairly straightforward process -- one can either choose a predictive prior based on a historical dataset, which need not be very large, or an uninformed prior in the absence of such data. A unique feature about any Bayesian approach is that the impact of the prior on the posterior predictive distribution diminishes as more Bayesian updates are made in the light of new data. This section presents details on how to compute posterior predictive distributions of task durations, using priors and likelihood functions in a real-time task planning or scheduling system.

Consider a planning or a scheduling system in which task duration estimates for tasks are needed. After determining an appropriate model for the task distribution -- and any models necessary for covariate estimations -- to bootstrap the system, set prior distributions for all model parameters in the hierarchy. Once data is observed, a Bayesian update is performed to obtain the set of posterior distributions. These distributions are then used as priors for the next Bayesian update, and are used to obtain the posterior predictive distribution for the task duration. As mentioned earlier, closed form solutions for the posterior distributions are generally not available, and often they are computed using numerical integration [15], MCMC [16] methods, or nested sampling techniques [17]. In this paper, we use the Metropolis Hastings algorithm to obtain MCMC samples of the posterior distribution over a set of parameters. The specific details of this algorithm are presented in Algorithm 3.

Posterior distributions of the parameters are used in computing the posterior predictive task duration distribution. A choice descriptive statistic (e.g., mean or median) of the resulting task duration distribution can be used to inform control decisions should point estimates be preferred to probability distributions. For example, typically, a precision parameter (or variance) of the posterior predictive distribution provides insight into "how good" a specific prediction is. In fact, one can make use of this information to make decisions on whether to incorporate a specific prediction value in task planning and scheduling.

Lastly, while designing the system, it is important to pay attention to the convergence and mixing properties of numerical integration algorithms (in this specific case MCMC). Failing to do so may result in model parameters converging to point distributions. As noted by Brown et al. [18], there are three conditions under which MCMC posterior parameter estimate might converge to a point distribution: 1) existence of multiple local peaks in the posterior will make it difficult for MCMC algorithm to traverse the space of parameters; 2) even if the posterior is single moded, MCMC does not mix well due to the existence of equal posterior density for a large regions of the posterior; 3) Overly informative priors favors unreasonable large branch lengths. In theory, these problems can be tackled by specifying compound Dirichlet priors for branch lengths. However, this can also be prevented by ensuring the standard deviation of the posterior doesn't converge to zero. In this work, we empirically determined lower bounds on the standard deviation of each parameter. If the standard deviation of any parameter's posterior distribution falls below the preset lower bound, the parameter is reset to have a Normal distribution with the same mean and a standard deviation above the lower bound, which was also empirically determined.

Algorithm 3 Real-Time Bayesian Inference for Task Duration Under Uncertainty

$C_t \leftarrow$ predicted (or observed) covariates at time t
$M \leftarrow$ most recent MCMC parameter samples
$\sigma_t \leftarrow$ lower threshold for standard deviation of M
$\sigma_r \leftarrow$ standard deviation to reset M
$P \leftarrow$ the variable to predict
$\hat{p}_t \leftarrow$ prediction for P generated at time t
set up priors for the parameters
while $t < \infty$ **do**
  if M is not `None` **then**
    $R \leftarrow$ samples of parameters from M, using C_t
    $S \leftarrow$ samples of P for each parameter sample in R using P's parametrized distribution
    $S_{med} \leftarrow$ median of each sublist in S
    $\hat{p}_t \leftarrow$ mean of S_{med}
  if new observation for P is available **then**
    $p_t \leftarrow$ new observation for P
    $K_M \leftarrow$ set of Kernel Density Estimates for each parameter in M
    $L \leftarrow$ likelihood of P, parametrized by the distributions of parameters K_M and covariates C_t
    $M \leftarrow$ samples of approximate posterior distribution of the parameters obtained with Metropolis-Hastings algorithm on previous M, using L and p_t
    for $M_i \leftarrow$ samples for parameter $i \in M$ **do**
      $\sigma_i \leftarrow$ standard deviation of M_i
      if $\sigma_i < \sigma_t$ **then**
        $\mu_i \leftarrow$ mean of M_i
        $M_i \leftarrow$ samples from $N(\mu_i, \sigma_r^2)$

It is important to note that this algorithm is used in a rolling fashion to make task duration predictions for each task in real-time. This is a major advantage of the framework because there is no need for a training dataset to learn the model parameters, since they are estimated online via Bayesian updates. As we will demonstrate in subsequent sections of this paper, this framework is able to generate highly predictive models of task durations that are resilient to non-stationary stochastic processes.

APPLICATION: BUS DWELL TIME PREDICTION

Background

Real-time optimization of the dynamic flow of vehicle traffic through a network of signalized intersections is an important practical problem. It is well known that the vehicle flows at signalized intersections constitute a non-stationary stochastic process, and optimal control of those flows is NP-hard [19]. To cope with the inefficiency of searching in an exponential planning search space, distributed online planning approaches are proposed for real-time signal control [20, 21].

For example, according to the Surtrac planning algorithm [21], at each planning cycle, each intersection constructs a prediction of the sequence of arriving vehicles from its local sensors and then constructs a "signal timing plan" (an allocation of green time to various approaches) in real-time that moves detected vehicles through the intersection in a way that minimizes cumulative wait time. As the intersection begins executing the plan, it also sends an expectation to its downstream neighbors of what traffic it expects to be sending their way, giving those intersections the "visibility" to plan over a longer horizon. Intersection plans are executed in rolling horizon fashion and the planning process repeats every couple of seconds.

Furthermore, all these optimization algorithms rely on a prediction of vehicle arrival times in the form of sequences of vehicle clusters that are detected along different approaches. They use an aggregate representation of approaching traffic flows as sequences of clusters (i.e., queues and platoons) and predict arrival time of these clusters strictly based on use of free flow speed. However, this aggregate representation does not distinguish between vehicle classes (e.g., passenger cars, transit vehicles like buses etc) that might have very different flow patterns and hence arrival times. For example, unlike passenger cars, transit vehicles make frequent stops (to pick up or drop off passengers) with uncertain dwell times. The presence of transit vehicles stopping on urban streets can also restrict or block other traffic on the road depending on stop locations. As a consequence, executable schedules generated by real-time online planning algorithms are disrupted resulting in inefficient traffic flows. These planning algorithms can generate better anticipatory schedules if they have reasonably good estimates of vehicle cluster arrival times. The framework discussed in the previous section can play an instrumental role in providing such estimates.

In principle, one can treat vehicle cluster arrival times as various tasks, where the goal is to predict task durations and quantify associated uncertainties. For the reasons mentioned above, accurate transit bus dwell time predictions play a vital role in achieving this goal. In that regard, the purpose of this application demonstration is to test the efficacy of the proposed Bayesian framework for task duration predictions (in this context bus dwell times).

Model Overview

Constructing a predictive bus dwell time distribution model involves three sub-tasks: 1) choosing the likelihood function for posterior updates; 2) choosing principal covariates that influence dwell time distributions; and 3) formalizing a dwell time model using information from the previous sub-tasks.

Likelihood Function For Posterior Updates

Consistent with the guidance provided in the framework, we used historical data for choosing a likelihood function. Specifically, we used the Port Authority of Allegheny County's (PAAC) Advanced Vehicle Location (AVL) weekday dataset. The data is chronologically ordered, and empirical CDFs based on every fifteen minutes of data are created. Dwell times in the APCC dataset are rounded to the nearest second. To address this, two different continuous empirical CDFs are generated using Gaussian, and Gamma KDE techniques. Next, using the same temporally sequential data six analytic distributions (Non-central F, Burr, Weibull, Beta, Log-normal, and Fisk or Log-logistic) are generated. Max-deviation scores are computed between each analytic distribution fit and each of the two empirical distributions. Based on MDT scores, we chose the Log-logistic (Fisk) distribution as the likelihood for the posterior updates.

Covariates For Dwell Times

In order to develop a dwell time model with covariates, several relationships were explored between covariate data and dwell time, such as the number of onboarding passengers (x_{on}), number of alighting passengers (x_{off}), and load of the bus (x_{load}), A clear positive correlation was found between first two covariates and dwell time, which were chosen as covariates in developing the predictive dwell time distribution model. A scatter plot demonstrating the relationship between the number of onboarding passengers and the dwell time is presented in Figure 1. Figure 2 demonstrates not only that more onboarding passengers corresponds to longer dwell times, but also that the variance of the dwell time increases as more passengers board.

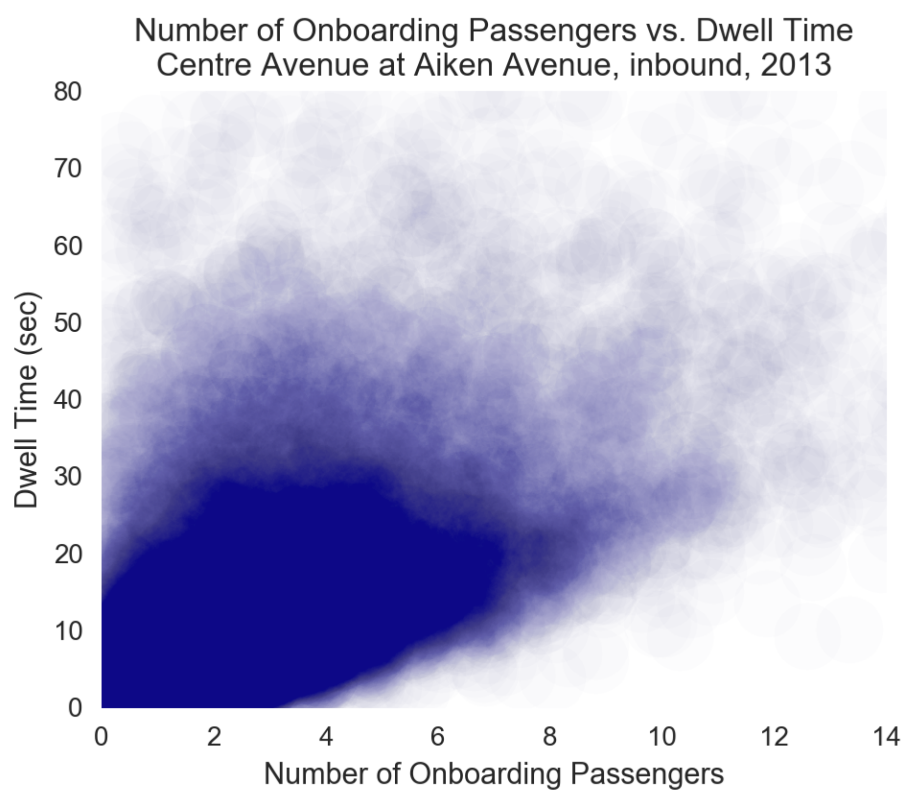

Figure 1: Scatterplot of # onboardings vs. dwell times

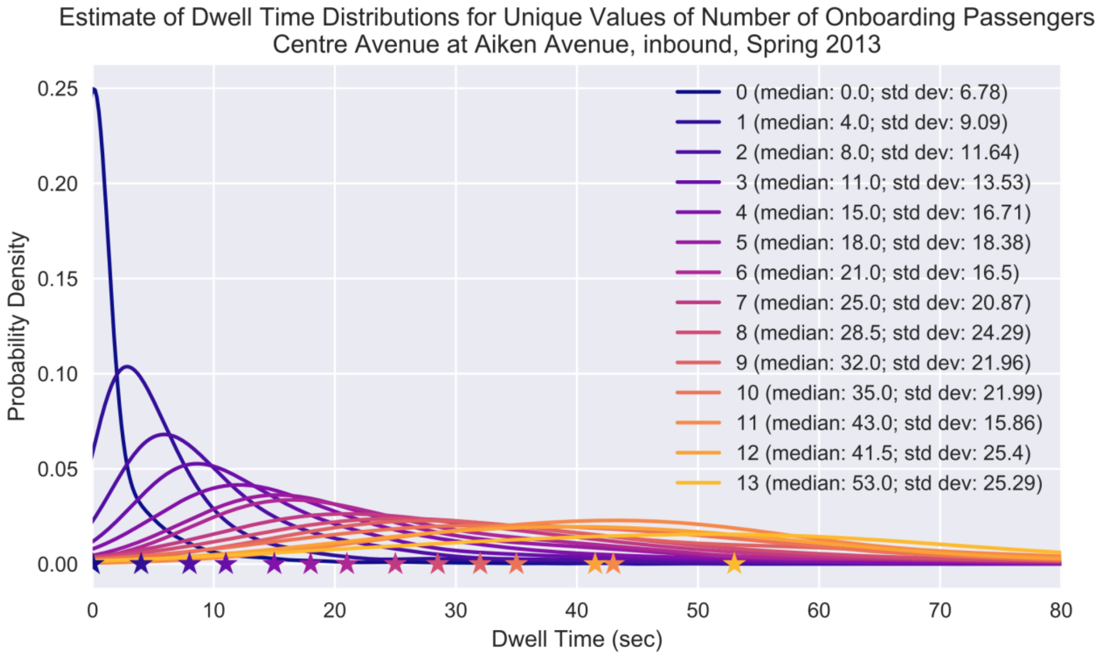

Figure 2: Conditional dwell time distributions for several numbers of onboarding passengers. Note that the variance is larger when more passengers board.

Dwell Time Model With Covariates

The following describes a Bayesian parametric model for bus dwell times using two covariates $\mathbf{x}_{\mathbf{on}}$ and $\mathbf{x}_{\mathbf{off}}$. Based on the analysis presented in the subsection on choosing the likelihood function, bus dwell time is modeled as a random variable **X** following a Log-Logistic (Fisk) distribution. Equivalently, bus dwell times **X** are distributed following the exponential of the Logistic distribution. Covariate parameters are introduced by parameterizing the **s** parameter, and the median of the Log-Logistic distribution. The exponential relationship between the Logistic and Log-Logistic distributions is used in this formulation. This parameterization is described below:

$$X = \exp(Y)$$

$$\text{Where } Y \sim \text{Logistic}(\mu, s)$$

$$\mu = \ln(\alpha) = \ln(\beta_\alpha^T x + \beta_0)$$

$$s = {}^{1}/_{\tau} = {}^{1}/_{(\beta_\tau^T x)}$$

$$\beta_\alpha = \begin{bmatrix}\beta_\alpha^{on} & \beta_\alpha^{off}\end{bmatrix}^T$$

$$\beta_\tau = \begin{bmatrix}\beta_\tau^{on} & \beta_\tau^{off}\end{bmatrix}^T$$

$$x = \begin{bmatrix}x_{on} & x_{off}\end{bmatrix}^T$$

At any given time, the belief of the two parameters **μ** and **s** describe current belief of bus dwell time distribution. In a real-time system with access to dwell time observations, belief of the parameter distributions is continuously updated in the light of new data. Bayes' Theorem offers a natural way to achieve such an update scheme. As only one observed dwell time **d** is considered during any Bayesian update, the likelihood function is given by

$$L(\mu, s|\ln(d)) = f(\ln(d), \mu, s)$$

where f is the probability density function of a Logistic distribution.

Before obtaining any posterior distributions to use as priors, we bootstrap the model using a Normal prior for each of the 4 covariate parameters: β_α^{on}, β_α^{off}, β_τ^{on}, β_τ^{off} and offset parameter β_0. Once a set of posterior distributions is obtained, the most recent posterior distributions are used as priors in the next Bayesian update. The Metropolis Hastings algorithm is employed to obtain MCMC samples of the posterior distributions for four covariate parameters and the offset parameter.

To make a dwell time prediction for an approaching bus, we observe values for covariates x_{on}, and x_{off}, and use posterior distributions of each β to determine the posterior predictive distribution of **X**.

This process is repeated in the light of new data, using the most recent posterior distributions of each β as priors in the next Bayesian update. The means of several model parameters are shown in Figures 3 and 4, where a real-time prediction scenario is simulated on historical data in a rolling fashion.

MODEL TESTING

We tested the efficacy of the proposed dwell time prediction model on bus dwell time data provided by the Port Authority of Allegheny County in Pittsburgh, Pennsylvania for the period from September 2012 to August 2014. While the dataset spans over two years, data from October 2012 is used to test the Bayesian model. We compared the results of the Bayesian model to those of a linear regression model for benchmarking purposes. We trained a linear regression model on September 2012 and tested on October 2012, which are good training and test datasets since it is widely accepted that seasonal trends in bus dwell time distributions are statistically similar [22] (also, readers interested in dwell time distribution models can find comprehensive reviews in [22]). Therefore, the linear regression model is not really put to the test. In principle, regression equations for September 2012 & October 2012 should look very similar, suggesting that predictions on the test dataset should be reasonably good. However, the main objective of this analysis is to evaluate the robustness of the proposed framework. In other words, the goal is to check whether the Bayesian model is able to predict dwell times without any training and how good those predictions are compared to predictions from a well-trained traditional model.

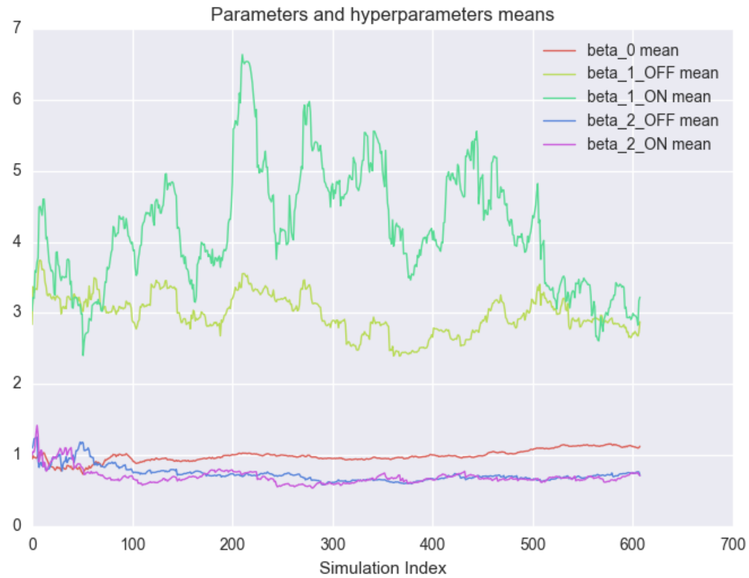

Figure 3: Means of model parameters throughout simulation. beta_1 corresponds to beta_alpha, beta_2 corresponds to beta_tau

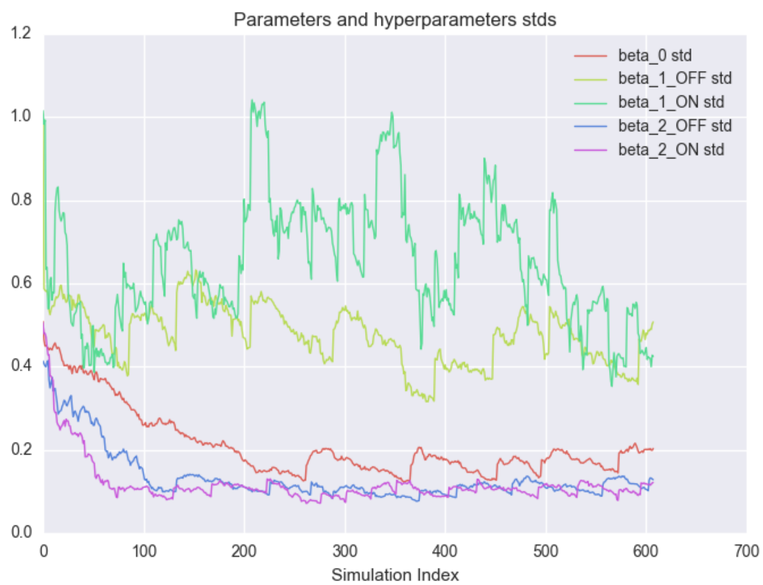

Figure 4: Standard deviations of model parameters throughout simulation. Note that the MCMC samples are reset when the standard deviation falls below a specified threshold.

With these objectives in mind, we tested the robustness of the Bayesian framework at twelve different bus stops in the East End region along Centre Avenue corridor in Pittsburgh, PA.

Cumulative Density Functions Of Dwell Times

Analyzing cumulative density functions (CDFs) of dwell times provides useful insights into the reliability (presence or absence of variance) of these distributions. From the standpoint of stochastic dominance, the distributions with curves furthest to the left have smaller variance in dwell time distributions and hence are more reliable.

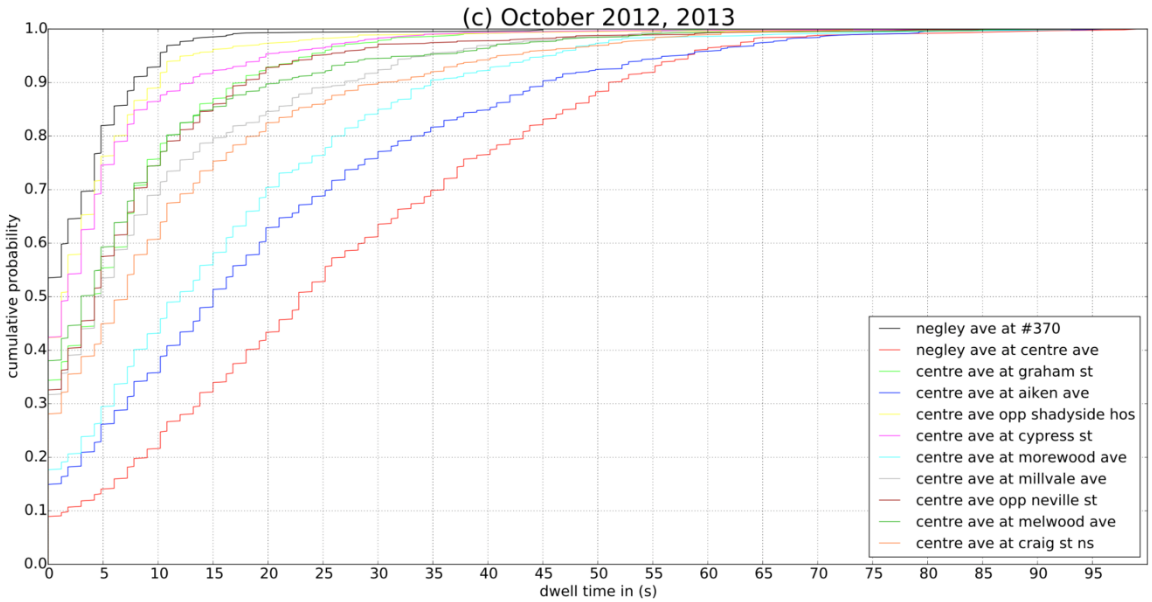

Figure 5: Cumulative density functions of dwell times

Figure 5 presents dwell time CDFs for test bus stops of interest. It can be seen that dwell time distributions have the largest variance at Negley Ave at Centre Ave (CDF in red), followed by Centre Ave at Aiken Ave (blue), Centre Ave at Morewood Ave (cyan), Centre Ave at Craig St NS (peach), and Centre Ave at Millvale (light grey). This information is useful because predicting dwell time distributions at these intersections is particularly hard due to their highly stochastic nature.

Model Performance

As mentioned earlier, the efficacy of the Bayesian model is evaluated on data from October 2012 and the results are benchmarked against those obtained from a linear regression model, which is trained at each bus stop on September 2012 data. The same Bayesian parametric model is applied to each of the bus stops, and we set Normal priors for each of the 4 covariate parameters and the offset parameter β_0. Covariate parameters are updated on an ex post facto basis, and dwell time predictions are made starting from the very first new data point onward.

We use the ability to predict dwell times within a small error threshold as a performance metric to evaluate the models. The rationale for choosing small error bounds is to account for the fact that these dwell time values are used by planning algorithms in real-time systems, so larger errors will generate schedules that are far from optimal. For this reason, the fraction of predictions within error bounds of [-5, 5] seconds is used as a performance metric. Effectively, this fraction represents the area under the error distribution density function within these tolerance bounds. This is a more informative metric in the context of traffic signal scheduling due to the importance of maximizing the proportion of very close predictions.

Table 1 summarizes performance of these two models. As can be seen, this table contains three sets of performance comparisons: 1) morning peak hour (7:00 - 10:00 AM); 2) evening peak hour (4:00 - 7:00 PM); and 3) the entire test dataset. This table has four columns: first column presents bus stop location information; second column presents fraction of dwell time predictions within the range of -5 and 0 seconds; third and fourth columns contain similar information but for ranges of [0, 5] and [-5, 5] seconds respectively. Lastly, each row contains results for a specific bus stop.

The following inferences can be drawn based on these results: First, the Bayesian predictive model performs at least as good as or better than the linear regression model. This is very encouraging to see as it validates the main philosophy behind the development of this framework, i.e., to develop a predictive probabilistic model for estimating task durations without making use of large training datasets. Second, for the scenarios in which dwell time distributions are highly stochastic (refer Fig 5), the Bayesian prediction model significantly outperforms the linear regression model (refer to results for Negley Ave at Centre Ave, Centre Ave at Aiken Ave, and Centre Ave at Craig St NS). Figure 6 demonstrates this trend for Negley Ave at Centre Ave - the Bayesian model has a much higher proportion of very close predictions than the linear regression error distribution. This again corroborates the hypothesis of quick adaptability of the Bayesian model. Third, in addition to dwell time estimates, the variance or precision parameter of the Bayesian model quantifies the uncertainty of each prediction.

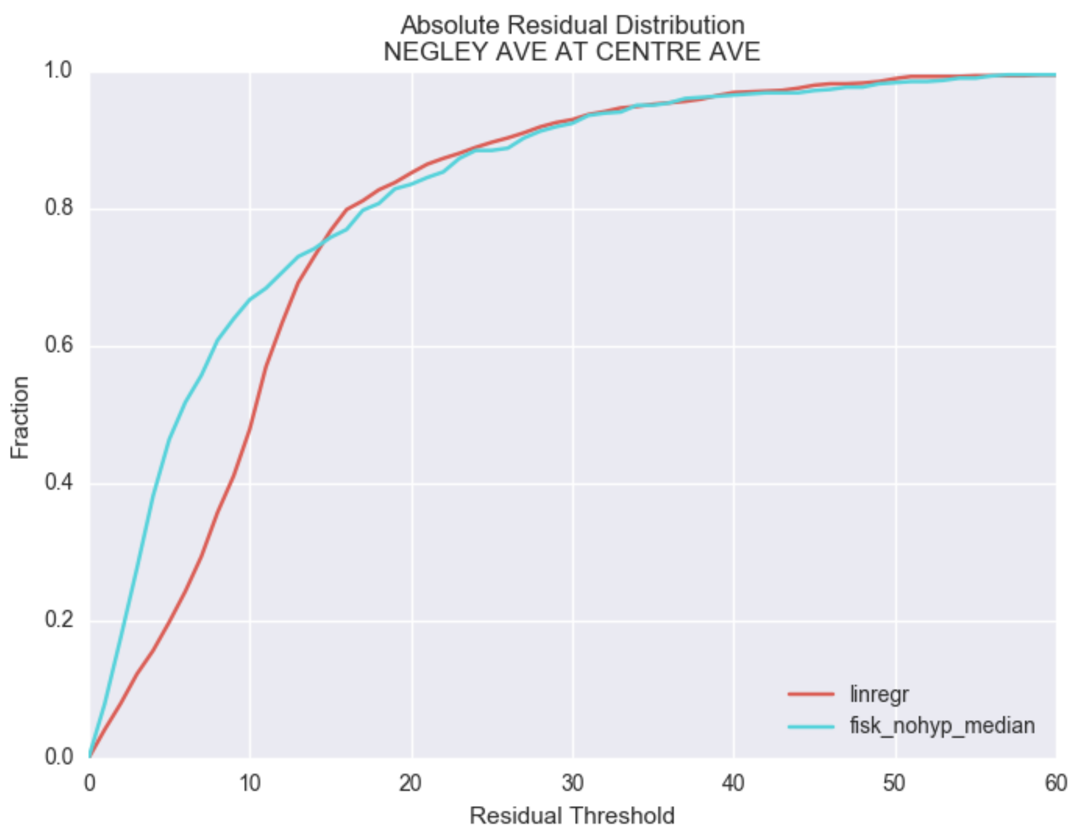

Figure 6: Fraction of absolute prediction error within a threshold for our framework vs. linear regression. Note that the Bayesian hierarchical model has a higher proportion of small errors.

Bus Stop		[-5, 0]		[0, 5]		[-5, 5]	
		L.R.	Fisk	L.R.	Fisk	L.R.	Fisk
Centre Ave at Aiken Ave	AM	0.11	0.22	0.31	0.29	0.42	0.51
	PM	0.11	0.20	0.40	0.44	0.51	0.64
	All	0.10	0.21	0.39	0.41	0.49	0.63
Negley Ave at Centre Ave	AM	0.13	0.18	0.13	0.33	0.27	0.51
	PM	0.08	0.14	0.09	0.33	0.16	0.47
	All	0.10	0.15	0.11	0.32	0.21	0.46
Negley Ave at #370	AM	0.20	0.33	0.59	0.52	0.79	0.84
	PM	0.13	0.21	0.66	0.66	0.79	0.87
	All	0.15	0.29	0.63	0.57	0.78	0.86
Centre Ave Opp Neville St	AM	0.23	0.31	0.45	0.48	0.69	0.79
	PM	0.18	0.26	0.55	0.47	0.73	0.73
	All	0.21	0.29	0.51	0.50	0.72	0.79
Centre Ave at Shadyside Hos	AM	0.25	0.36	0.54	0.45	0.79	0.80
	PM	0.22	0.27	0.48	0.48	0.70	0.75
	All	0.20	0.30	0.52	0.46	0.73	0.76
Centre Ave at Morewood Ave	AM	0.16	0.25	0.38	0.35	0.54	0.61
	PM	0.16	0.37	0.58	0.44	0.73	0.81
	All	0.15	0.28	0.52	0.45	0.67	0.73
Centre Ave at Millvale Ave	AM	0.14	0.25	0.41	0.45	0.54	0.71
	PM	0.18	0.35	0.50	0.44	0.68	0.79
	All	0.13	0.28	0.51	0.50	0.65	0.78
Centre Ave at Melwood Ave	AM	0.21	0.25	0.46	0.51	0.67	0.76
	PM	0.23	0.30	0.54	0.50	0.77	0.79
	All	0.20	0.28	0.54	0.52	0.74	0.80
Centre Ave at Graham St	AM	0.17	0.34	0.57	0.44	0.73	0.78
	PM	0.24	0.34	0.53	0.42	0.76	0.76
	All	0.19	0.31	0.59	0.48	0.78	0.80
Centre Ave at Cypress St	AM	0.19	0.30	0.55	0.47	0.74	0.77
	PM	0.09	0.26	0.53	0.42	0.62	0.68
	All	0.16	0.26	0.54	0.46	0.70	0.73
Centre Ave at Craig St NS	AM	0.07	0.17	0.25	0.32	0.32	0.49
	PM	0.15	0.13	0.26	0.30	0.42	0.43
	All	0.11	0.17	0.26	0.34	0.37	0.51
Centre Ave Opp Shadyside Hos	AM	0.22	0.32	0.60	0.51	0.82	0.82
	PM	0.21	0.28	0.54	0.50	0.76	0.78
	All	0.21	0.29	0.57	0.49	0.79	0.78

TABLE I: Model Performance Comparisons

Hierarchical Bayesian Model

To demonstrate the ideas of hierarchical model, a variant of dwell time estimation model is considered. This model takes two input covariates: 1) estimated value of number of onboardings ($\tilde{x}_{on}$), and 2) observed value of number of alightings (x_{off}). Arrival rate of passengers at a bus stop can be modeled as a doubly stochastic Poisson process, and we developed a Bayesian model to estimate these arrival rates. This model uses predicted arrival rate and known bus headway in estimating $\tilde{x}_{on}$. The model details are presented below.

Let Y_i is the number of passengers boarding onto the bus during a bus arrival event i. The arrival rate of passengers at a bus stop is modeled using λ parameter of a Poisson distribution. For the purpose of Bayesian updates, the posterior for λ represented by p(λ | y) is derived as:

$$p(y/\lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} \propto \lambda^{ny} e^{-n\lambda}$$

This is the kernel of a Gamma distribution. Therefore, if $\lambda \sim Ga(\alpha, \beta)$, then

$$p(\lambda/y) \propto p(y/\lambda) p(\lambda)$$

$$p(\lambda/y) \propto \lambda^{n\tilde{y}} e^{-n\lambda} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$p(\lambda/y) = \lambda^{\alpha+n\tilde{y}-1} e^{-(\beta+n)\lambda}$$

$$\lambda | y \sim Ga(\alpha + n\tilde{y}, \beta + n)$$

where β is number of prior observations; α is the sum of previous arrival rates.

A non-informative prior such as Jeffreys' prior is used to bootstrap the system. So $p(\lambda) \propto J(\lambda)^{1/2}$ where $J(\lambda)$ is the Fisher information, which is the negative expectation of the second derivative of the log likelihood.

$$\log p(y|\lambda) = -\log(y!) + y \log(\lambda) - \lambda \text{ (log likelihood)}$$

The second derivative of the above function is equal to $-y/\lambda^2$

$$J(\lambda) = -E\left[-\frac{y}{\lambda^2}\middle|\lambda\right] = \frac{1}{\lambda}$$

$$J(\lambda)^{1/2} = \frac{1}{\sqrt{\lambda}}$$

The previous equation can be treated as Ga(1/2, 0). Note that this is an improper Gamma distribution, but it is acceptable for the purpose of Bayesian updates.

In order obtain a posterior arrival rate distribution via a Bayesian update, a list of observed arrival rates are maintained, which are defined by the number of onboardings divided by the headway. Once a new observation (headway and onboardings) is made, the arrival rate is computed and appended to the list. A new value for α is calculated as sum of the recent β arrival rate observations, where β is an integer that should be empirically found to maximize prediction accuracy. An onboarding prediction for an approaching bus is made by multiplying a point estimate of the posterior arrival rate distribution (e.g., mean, median) with the headway. Here the headway information can be obtained from published bus time tables.

The hierarchical model is tested at five out of twelve intersections, and results are summarized in Table 2. The results are benchmarked against any traditional learning model, as the main idea is to demonstrate details of the hierarchical Bayesian framework.

Bus Stop		[-5,5]
Centre Ave at Aiken Ave	AM	0.36
	PM	0.60
	All	0.50
Negley Ave at Centre Ave	AM	0.43
	PM	0.35
	All	0.42
Negley Ave at #370	AM	0.82
	PM	0.81
	All	0.83
Centre Ave at Craig St NS	AM	0.49
	PM	0.40
	All	0.47
Centre Ave at Shadyside Hos	AM	0.72
	PM	0.78
	All	0.67

Table 2: Hierarchical Fisk Model

Conclusions And Future Work

This paper presents a hierarchical Bayesian predictive probabilistic model for task duration predictions in real-time systems. The framework is computationally efficient, reduces the problem of overfitting, and requires little or no training to start producing good predictions. Furthermore, unlike traditional learning models, the proposed framework effectively addresses uncertainty by delivering a confidence in the prediction through the posterior predictive distribution, rather than simply supplying a point estimate.

The ideas presented in the framework are tested in the context of predicting dwell time distributions of a transit buses in urban networks. Specifically, a Bayesian parametric model for bus dwell times was created using two covariates, x_{on} , and x_{off}. The efficacy of this model is tested at twelve different bus stops in the East end region of Pittsburgh, PA on real-world bus dwell time data. The results of the model are benchmarked against those obtained from a linear regression model. The results demonstrate that the Bayesian model is able to perform at least as good as, and in most instances far better than traditional learning models.

Finally, to demonstrate the ideas of hierarchical models, a new dwell time estimation model was considered. The input parameter x_{on} was estimated, whereas the other parameter x_{off}. was observed. Model details are presented for estimating $\tilde{x}_{on}$. The hierarchical model was tested at the twelve intersections and the results do validate the usefulness of the framework.

We envision two future directions to this research: First, we are interested in integrating the bus dwell time model into an online planning algorithm like Surtrac to investigate the system performance improvements. Second, we want to investigate the efficacy of this framework in other domains of planning & scheduling.

REFERENCES

1. R. Dechter, I. Meiri, and J. Pearl, "Temporal constraint networks,"Artificial Intelligence, vol. 49, no. 1-3, pp. 61–95, May 1991..

2. T. Vidal and G. M., "Dealing with uncertain durations in temporal constraint networks dedicated to planning," in Proceedings 12th European Conference on Artificial Intelligence (ECAI-1996), 1996, pp. 48–54.

3. P. Morris, M. Muscettola, and T. Vidal, "Dynamic control of plans with temporal uncertainty," in Proceedings 17th international joint conference on Artificial Intelligence, Seattle, WA, August 2001, pp. 494–499.

4. N. Policella, A. Cesta, A. Oddi, and S. Smith, "Solve-and-robustify: Synthesizing partial order schedules by chaining," Journal of Scheduling, vol. 12, no. 3, 2009.

5. M. Drummond, J. Bresina, and K. Swanson, "Just in-case scheduling," in Proceedings AAAI-94, 1994.

6. H. Younes and R. Simmons, "Policy generation for continuous time stochastic domains with concurrency," in Proceedings ICAPS-04, Whistler, Canada, 2004.

7. I. Little, D. Aberdeen, and T. S., "Prottle: A probabilistic temporal planner," in Proceedings AAAI-05, 2005.

8. Mausam and D. Weld, "Planning with durative actions in uncertain domains," Journal of Artificial Intelligence Research, vol. 31, pp. 33–82, 2008.

9. J. Brooks, A. Reed, E.and Gruver, and J. Boerkoel, "Robustness in probabilistic temporal planning," in Proceedings AAAI-15, 2015, p.3239–3246.

10. S. F. Smith, A. Gallagher, T. L. Zimmerman, L. Barbulescu, and Z. B.Rubinstein, Distributed management of flexible times schedules," in Proceedings 6th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 07), Honolulu Hawaii, May 2007.

11. S. Yoon, A. Fern, R. Givan, and S. Kambhampati, "Probabilistic planning via determinization in hindsight," in Proceedings 23rd AAAI Conference on Artificial Intelligence, 2008.

12. I. K. Isukapati and G. F. List, “Synthesizing route travel time distributions considering spatial dependencies,” in Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on. IEEE, 2016, pp. 2143–2149.

13. R. Wilcox, “Kolmogorov–smirnov test,” Encyclopedia of biostatistics, 2005.

14. P. J. Moreno, P. P. Ho, and N. Vasconcelos, “A kullback-leibler divergence based kernel for svm classification in multimedia applications,” in Advances in neural information processing systems, 2004, pp. 1385–1392.

15. L. Tierney and J. B. Kadane, “Accurate approximations for posterior moments and marginal densities,” Journal of the american statistical association, vol. 81, no. 393, pp. 82–86, 1986.

16. S. Chib and E. Greenberg, “Understanding the metropolis-hastings algorithm,” The american statistician, vol. 49, no. 4, pp. 327–335, 1995.

17. J. Skilling et al., “Nested sampling for general bayesian computation,”Bayesian analysis, vol. 1, no. 4, pp. 833–859, 2006.

18. J. M. Brown, S. M. Hedtke, A. R. Lemmon, and E. M. Lemmon, “When trees grow too long: investigating the causes of highly inaccurate bayesian branch-length estimates,” Systematic Biology, vol. 59, no. 2, pp.145–161, 2009.

19. C. H. Papadimitriou and J. N. Tsitsiklis, “The complexity of optimal queuing network control,” Mathematics of Operations Research, vol. 24, no. 2, pp. 293–305, 1999.

20. S. Sen and K. L. Head, “Controlled optimization of phases at an intersection,” Transportation science, vol. 31, no. 1, pp. 5–17, 1997.

21. S. F. Smith, G. J. Barlow, X.-F. Xie, and Z. B. Rubinstein, “Smart urban signal networks: Initial application of the surtrac adaptive traffic signal control system.” in ICAPS, 2013.

22. Isukapati, I. K., Rudová, H., Barlow, G. J., & Smith, S. F. (2017). Analysis of Trends in Data on Transit Bus Dwell Times. Transportation Research Record: Journal of the Transportation Research Board, (2619), 64-74.