Hierarchical Bayesian Framework For Bus Dwell Time Prediction

Isaac K Isukapati[1], Conor Igoe[1], Eli Bronstein[2], Viraj Parimi[1], and Stephen F Smith[1]
[1] The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA
[2]Department of Electrical Engineering & Computer Science, University of California, Berkeley, CA 94720 USA

In many applications, uncertainty regarding the duration of activities complicates the generation of accurate plans and schedules. Such is the case for the problem considered in this paper - predicting the arrival times of buses at signalized intersections. Direct vehicle-to-infrastructure communication of location, speed and heading information offers unprecedented opportunities for real-time optimization of traffic signal timing plans, but to be useful bus arrival time prediction must reliably account for bus dwell time at near-side bus stops. To address this problem, we propose a novel, Bayesian hierarchical approach for constructing bus dwell time duration distributions from historical data. 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, all of which are important in the bus dwell time prediction context. The effectiveness of this approach is demonstrated using historical data provided by a local transit authority on bus dwell times at urban bus stops. Our results show that the dwell time distributions generated by our approach yield significantly more accurate predictions than those generated by both standard regression techniques and a more data intensive deep learning approach.

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

I. INTRODUCTION

MANY practical planning and scheduling problems are complicated by the durational 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 for example, and hence may require multiple attempts before the larger plan in which the task is embedded can move forward. Alternatively, a vehicle traveling from a given pickup location to a given drop-off location may have several different routes to choose from, with each route having variable duration and being dependent on current traffic conditions. Effective planning and scheduling in such circumstances requires the ability to accurately characterize this uncertainty.

In this paper, we consider this task duration modeling challenge in a particular setting, that of predicting bus dwell times at bus stops in urban road networks. Reliable prediction of bus dwell times at near-side bus stops is crucial for determining bus arrival times at signalized intersections, which in turn opens new opportunities for real-time optimization of urban traffic flows. We focus on developing probabilistic dwell time duration models for individual bus stops from historical data, which can then be sampled for purposes of real-time prediction. We propose a novel Bayesian hierarchical approach to constructing probability models that offers several advantages over traditional statistical learning techniques in this application context, including the ability to start making accurate predictions with only minimal past data, the ability to provide robustness in the midst of a stochastic and noisy underlying system, and the ability to deliver measurable confidence in predictions. To demonstrate the efficacy of the approach, we present the results of experiments performed using historical data provided by a local transit authority. Our results show that significantly more accurate dwell time distributions can be derived from far less data than is possible with either standard linear regression methods, or more contemporary deep learning techniques.

Manuscript received June 14, 2018. Corresponding author: I. Isukapati (email: isaack@cs.cmu.edu).

The remainder of the paper is organized as follows. We first motivate and provide background information on the bus dwell time prediction problem. Next, we describe our Bayesian hierarchical modeling methodology for constructing bus dwell time models. This is followed by a presentation of our empirical analysis of its effectiveness in comparison to other candidate approaches. We then briefly discuss the potential broader applicability of the work to the general problem of generating task models for planning and scheduling systems. Finally, we summarize the main contributions of the paper and briefly indicate our future research directions.

II. BUS DWELL TIME PREDICTION PROBLEM

As indicated above, our interest in the accurate prediction of bus dwell times is motivated by the opportunity that it would provide to improve the real-time dynamic flow of vehicle traffic through a network of signalized intersections. It is well known that vehicle flows at signalized intersections constitute a non-stationary stochastic process, and optimal control of those flows is NP-hard [1]. Historically, this problem has been approached by estimating average flow conditions and developing a fixed signal timing plan (i.e., a fixed ordering and allocation of green time to various approaches at each intersection) offline that optimizes for these average conditions. However, advances in distributed computing over the past two decades have enabled the development of online planning approaches that produce signal timing plans in real-time that match the actual traffic on the road. [2, 3]

The Surtrac planning algorithm [3] provides a representative example of this online planning approach to traffic signal control. At the beginning of each planning cycle, each intersection independently senses the traffic approaching in all directions and constructs a prediction of when all sensed vehicles will reach the intersection. This prediction is then interpreted as a special type of single machine scheduling problem, and solved is to produce a signal timing plan that minimizes the cumulative wait time of all vehicles. As each intersection begins executing its plan, it also communicates 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 a rolling horizon fashion and the planning cycle at each intersection repeats every second.

To manage complexity, the arrival time prediction utilized by each intersection aggregates individual sensed vehicles that are approaching from various directions into sequences of clusters (i.e., queues and platoons) based on proximity. Since contemporary vehicle detection devices are not typically capable of providing mode information in real-time, clusters in this aggregate representation treat all constituent vehicles similarly. Arrival time prediction of clusters is based on a single "free flow" speed parameter.

With the advent of technologies that support direct vehicle-to-infrastructure (V2I) communication, it becomes straightforward to detect vehicle mode in real-time, and distinguish between different classes of vehicles (e.g., passenger vehicles, buses, bicyclists, etc.) that can have very different flow patterns, and resulting 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. Historically, Transit Signal Priority (TSP) systems have been introduced to streamline and expedite bus movements [4–8]. However, as pointed out by Isukapati et al. [9], by giving unconditional priority to transit vehicles, these traffic control strategies fail to optimize the overall traffic flow.

With V2I communication, it will be possible to produce a more accurate prediction of when buses will arrive at the intersection. For example, we would expect knowledge of an approaching bus with an intervening bus stop to trigger a disaggregation of its enclosing cluster, since the bus will block some trailing vehicles while it is stopped and hence should be the head of its cluster. A bus stop dwell time model can then be used to introduce expected cluster delay and propagate the effects to any additional blocked clusters that may reach the queue behind the bus before it leaves the bus stop.

The biggest source of uncertainty in this process is reliably predicting bus dwell times. Isukapati et al. [9], summarized the statistical characteristics of bus dwell times: 1) they vary considerably from stop to stop, 2) in addition to seasonal trends, the variance in dwell times over any interval is significant (so averages are not useful), and 3) given the huge variance in dwell time distributions, any predictive bus dwell time model needs to learn quickly and update distribution continuously to generate useful information in the context of real-time signal control. Consistent with these requirements, we propose a Bayesian hierarchical framework to predict bus dwell times in real-time. Previous work in predicting bus dwell times, which has focused on advance planning issues such as determination of bus schedules, has relied on linear regression prediction methods [10–13]. Furthermore, methodologies like KNNs [14, 15], random forests [16], and deep learning [17–23] are widely applied in time series forecasting Accordingly, our experimental analysis in Section V uses well trained linear regression, online weighted least squares, and a deep learning based LSTM approach for performance benchmark.

III. Bayesian Hierarchical Framework

State-of-the-art adaptive planning systems for real-time traffic control employ optimization models to decide how to allocate scare resources among tasks for optimal performance. These systems typically assume that current unfinished tasks have deterministic completion times rather than explicitly taking task duration uncertainty into account. Then, to account for dynamic behavior, optimization models are re-run upon the discovery of new information to generate updated optimal plans. These approaches tend to be reactive rather than proactive however, and are not likely to be effective for real-time traffic control. To generate signal timing plans that effectively optimize overall traffic flow in the presence of buses, it is necessary to utilize more informed models of bus dwell time duration that proactively quantify the uncertainty.

Given this goal, our approach is to utilize the availability of real-time (or near real-time) covariate task duration data (e.g. variables that influence bus dwell times such as the number of onboarding and alighting passengers) to produce more accurate duration models for bus dwell times at specific bus stops. 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 available data, necessitating a robust approach that is not prone to overfitting. Third, the available training datasets may be small, making models with many 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 a real-time decision-making system must be computationally efficient. Finally, being able to interpret the model and understand the structure of interactions of the variables is always an important requirement. In the following sub-sections, we introduce a Bayesian hierarchical framework that meets these requirements.

A. 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 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 continually updating its parameters 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 its *hierarchical* nature – the predictions of a "lower" model can be fed in as inputs to a "higher" model. For example, consider a task duration model with three input covariates $[x_1, x_2, x_3]$. As illustrated in Fig 1, only x_1, x_3 are directly observed, and x_2 is estimated by a model at different layer, and the estimated value of x_2 in turn, feeds as an input for the main task duration model. This concept is illustrated in the application section of this paper.

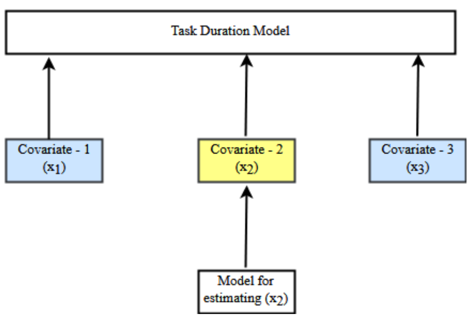

Fig. 1: Hierarchical Bayesian Framework

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

B. Selecting The Likelihood Function For Task Duration

The first step is to find an analytic distribution that best describes the empirical task duration distributions. Although, in principle, one could choose a distribution commonly used to model task durations, such as the log-logistic distribution, it is important to select the distribution that best matches historical data. 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 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. The next step is to statistically analyze similarities between the empirical CDF and each of the analytic distributions using the Maximum Deviation Test (MDT) [24].

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 $(t_l, t_u) \leftarrow (t_u, t_u + \eta)$

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 measures the statistical similarity between the empirical task duration distribution and each of the analytic distributions using MDT scores. The MDT score is defined as the number of percentile values in an analytic CDF (F_i') that are within a user-defined threshold of the empirical CDF (F). The analytic distribution with the highest MDT score (s_{max}) is statistically most similar to the empirical distribution. Pseudo-code for the methodology is given in Algorithm 2.

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}

Most non-parametric tests, such as the Kalmagorov-Smirnov (KS) test [25], 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 of 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 [26].

Note that if covariate data is not available a priori for prediction, this process can be also be used to determine an appropriate analytical distribution to estimate covariates. For example, while selecting likelihood function for dwell time distributions, we considered the six analytical distributions due to their common usage in survival analysis: Non-central F, Burr, Weibull, Beta, Log-normal, and Fisk (Log-logistic). We used small samples of historical data and MDT test to identify which of the six analytical distributions best explain the data.

C. Setting Priors And Generating Predictions

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 parameterized 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, parameterized 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 $N(\mu_i, \sigma_r^2)$

The next step after choosing the likelihood function(s) (the output of algorithm 1) 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. If the distribution parameters are expressed, for example, as a linear combination of input covariates, then prior distributions for each weight must be chosen. This is a fairly straightforward process – one can either choose a predictive prior based on a historical dataset 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.

Once the task duration analytic distribution and model parameter priors are chosen offline, the model can be deployed.Observed data is used to perform an online Bayesian update and obtain the posterior distribution over the model parameters. 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 [27], MCMC [28] methods, or nested sampling techniques [29]. In this paper, we use the Metropolis Hastings algorithm to obtain MCMC samples of the posterior distributions. The specific details of this algorithm are presented in Algorithm 3. Posterior distributions of the model parameters are used in computing the posterior 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. Moreover, 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 case MCMC). Failing to do so may result in model parameters converging to point distributions. As noted by Brown et al. [30], 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 mode, 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 does not converge to zero. In this work, we empirically determined lower bounds on the standard deviation of each parameter distribution. If the standard deviation of any parameter's posterior distribution falls below this lower bound, the parameter is reset to have a Normal distribution with the same mean and a standard deviation above the lower bound.

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. Thus, 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.

IV. Experimental Analysis

A. 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.

B. Likelihood Function For Posterior Updates

Consistent with the guidance provided in the framework (algorithms 1 & 2), 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 for the period from September 2012 to August 2014 for two major bus routes – 71A and 71C. 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 (as mentioned earlier, we choose these six analytic distributions due to their common usage in survival analysis). 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.

C. 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 Fig 2. Fig 3 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.

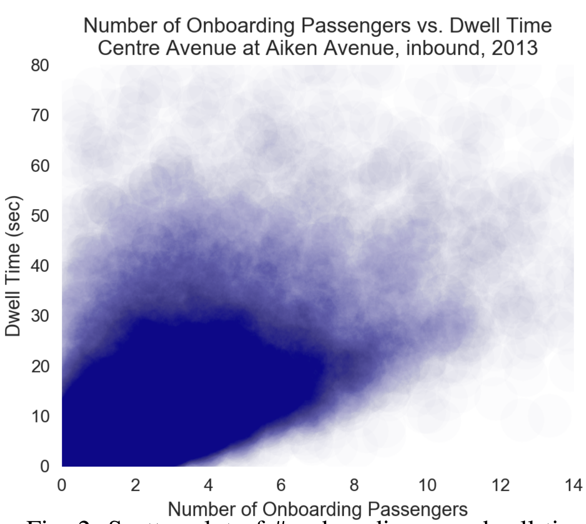

Fig. 2: Scatter plot of # onboardings vs. dwell times

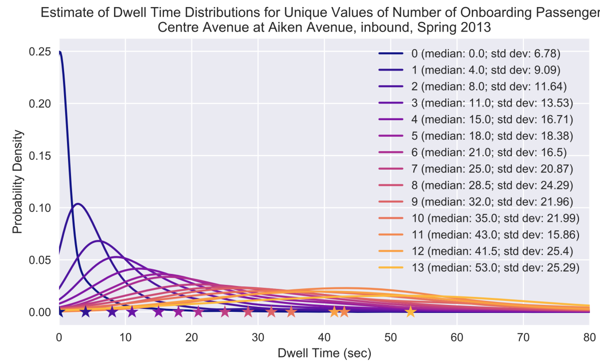

Fig. 3: Conditional dwell time distributions for several numbers of onboarding passengers. Note that the variance is larger when more passengers board.

D. Dwell Time Model With Covariates

The following describes a Bayesian parametric model for bus dwell times using two covariates x_{on} and 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:

$$\begin{aligned}
X &= \exp(Y) \qquad (\text{where } Y \sim \text{Logistic}(\mu, s)) \\
\mu &= \ln(\alpha) = \ln(\boldsymbol{\beta}_{\boldsymbol{\alpha}}^T \boldsymbol{x} + \beta_0) \\
s &= 1/\tau = 1/(\boldsymbol{\beta}_{\boldsymbol{\tau}}^T \boldsymbol{x}) \\
\boldsymbol{\beta}_{\boldsymbol{\alpha}} &= \begin{bmatrix} \beta_\alpha^{\text{on}} & \beta_\alpha^{\text{off}} \end{bmatrix}^T \\
\boldsymbol{\beta}_{\boldsymbol{\tau}} &= \begin{bmatrix} \beta_\tau^{\text{on}} & \beta_\tau^{\text{off}} \end{bmatrix}^T \\
\boldsymbol{x} &= \begin{bmatrix} x_{\text{on}} & x_{\text{off}} \end{bmatrix}^T
\end{aligned}$$

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}, $\beta_\alpha^{\text{off}}$, 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 and standard deviations of several model parameters are shown in Fig 4 and 5, where a real-time prediction scenario is simulated on historical data in a rolling fashion.

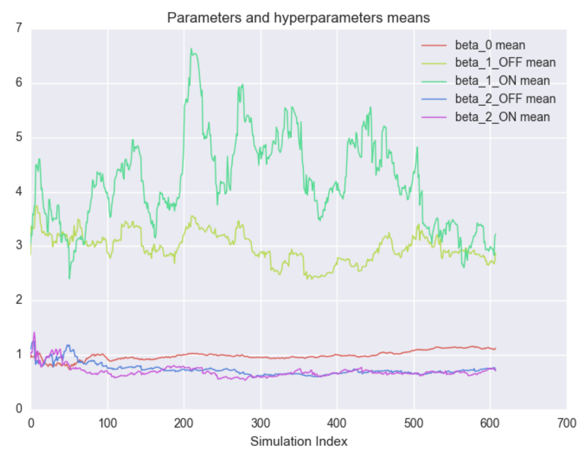

Fig. 4: Means of model parameters throughout simulation. beta_1 corresponds to β_α, beta_2 corresponds to β_τ

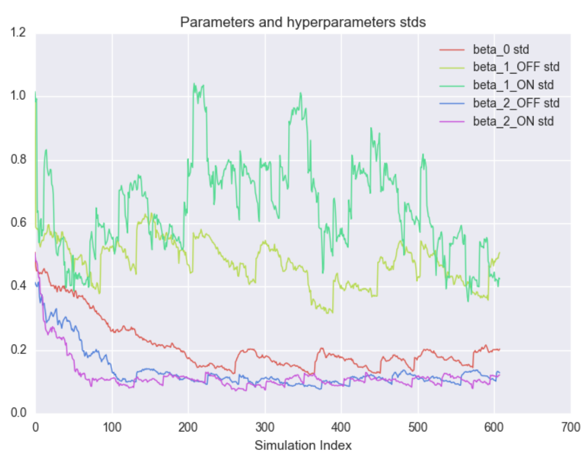

Fig. 5: Standard deviations of model parameters throughout simulation. Note that the MCMC samples are reset when the standard deviation falls below a specified threshold.

V. Model Testing

The efficacy of the proposed dwell time prediction model was tested 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 bench-marking 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 [9] (also, readers interested in dwell time distribution models can find comprehensive reviews in [9]). 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, the robustness of the Bayesian framework was evaluated at twelve different bus stops in the East End region along Centre Avenue corridor in Pittsburgh, PA.

A. 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.

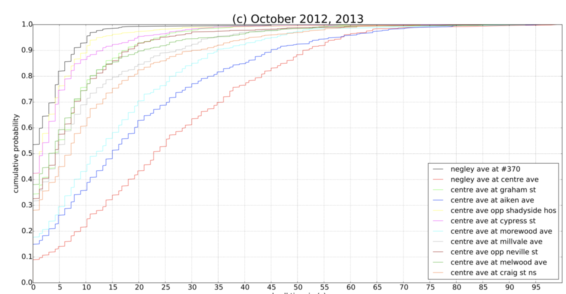

Fig. 6: Cumulative density functions of dwell times

Fig 6 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.

B. Model Performance

As mentioned earlier, the efficacy of the Bayesian model is evaluated on data from October 2012. The results are bench-marked against well trained linear regression, online least squares, and LSTM models trained offline 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

TABLE I: Model Performance Comparisons

Bus Stop		[-5, 0]				[0, 5]				[-5, 5]			
		L.R	Fisk	O.L.S	LSTM	L.R	Fisk	O.L.S	LSTM	L.R	Fisk	O.L.S	LSTM
Centre Ave at Aiken Ave	AM	0.11	0.22	0.03	0.08	0.31	0.29	0.38	0.11	0.42	**0.51**	0.42	0.19
	PM	0.11	0.20	0.06	0.39	0.40	0.44	0.32	0.17	0.51	**0.64**	0.38	0.22
	All	0.10	0.21	0.02	0.08	0.39	0.41	0.37	0.13	0.49	**0.63**	0.39	0.21
Negley Ave at Centre Ave	AM	0.13	0.18	0.02	0.02	0.13	0.33	0.16	0.05	0.27	**0.51**	0.19	0.07
	PM	0.08	0.14	0.05	0.08	0.09	0.33	0.17	0.10	0.16	**0.47**	0.22	0.18
	All	0.10	0.15	0.02	0.07	0.11	0.32	0.16	0.08	0.21	**0.46**	0.17	0.15
Negley Ave at # 370	AM	0.20	0.33	0.07	0.26	0.59	0.52	0.61	0.57	0.79	**0.84**	0.68	0.83
	PM	0.13	0.21	0.07	0.60	0.66	0.66	0.78	0.32	0.79	0.87	0.85	**0.92**
	All	0.15	0.29	0.02	0.19	0.63	0.57	0.71	0.66	0.78	**0.86**	0.73	0.85
Centre Ave Opp Neville St	AM	0.23	0.31	0.04	0.16	0.45	0.48	0.35	0.22	0.69	**0.79**	0.39	0.38
	PM	0.18	0.26	0.20	0.07	0.55	0.47	0.60	0.37	0.73	0.73	**0.80**	0.44
	All	0.21	0.29	0.02	0.12	0.51	0.50	0.50	0.34	0.72	**0.79**	0.52	0.46
Centre Ave at Shadyside Hos	AM	0.25	0.36	0.03	0.11	0.54	0.45	0.62	0.45	0.79	**0.80**	0.65	0.57
	PM	0.22	0.27	0.15	0.22	0.48	0.48	0.69	0.23	0.70	0.75	**0.84**	0.46
	All	0.20	0.30	0.03	0.15	0.52	0.46	0.53	0.28	0.73	**0.76**	0.56	0.43
Centre Ave at Morewood Ave	AM	0.16	0.25	0.05	0.08	0.38	0.35	0.41	0.12	0.54	**0.61**	0.46	0.20
	PM	0.15	0.37	0.11	0.12	0.58	0.44	0.44	0.35	0.73	**0.81**	0.55	0.47
	All	0.15	0.28	0.07	0.10	0.52	0.45	0.43	0.18	0.67	**0.73**	0.50	0.28
Center Ave at Millvale Ave	AM	0.14	0.25	0.16	0.10	0.41	0.45	0.49	0.18	0.54	**0.71**	0.65	0.27
	PM	0.18	0.35	0.25	0.16	0.50	0.44	0.53	0.25	0.68	**0.79**	0.79	0.41
	All	0.13	0.28	0.03	0.11	0.51	0.50	0.39	0.17	0.65	**0.78**	0.41	0.28
Centre Ave at Melwood Ave	AM	0.21	0.25	0.02	0.12	0.46	0.51	0.58	0.23	0.67	**0.76**	0.60	0.35
	PM	0.23	0.30	0.15	0.14	0.54	0.50	0.54	0.33	0.77	**0.79**	0.70	0.47
	All	0.20	0.28	0.02	0.11	0.54	0.52	0.54	0.24	0.74	**0.80**	0.56	0.35
Centre Ave at Graham St	AM	0.17	0.34	0.07	0.16	0.57	0.44	0.62	0.31	0.73	**0.78**	0.69	0.47
	PM	0.24	0.34	0.15	0.21	0.53	0.42	0.48	0.26	0.76	**0.76**	0.64	0.47
	All	0.19	0.31	0.02	0.16	0.59	0.48	0.53	0.25	0.78	**0.80**	0.55	0.41
Centre Ave at Cypress St	AM	0.19	0.30	0.09	0.07	0.55	0.47	0.59	0.30	0.74	**0.77**	0.68	0.37
	PM	0.09	0.26	0.15	0.24	0.53	0.42	0.65	0.24	0.62	0.68	**0.80**	0.47
	All	0.16	0.26	0.11	0.10	0.54	0.46	0.68	0.22	0.70	0.73	**0.79**	0.31
Centre Ave at Craig St NS	AM	0.07	0.17	0.03	0.11	0.25	0.32	0.30	0.16	0.32	**0.49**	0.33	0.28
	PM	0.15	0.13	0.18	0.06	0.26	0.30	0.38	0.18	0.42	0.43	**0.55**	0.24
	All	0.11	0.17	0.02	0.08	0.26	0.34	0.33	0.15	0.37	**0.51**	0.34	0.23
Centre Ave Opp Shadyside Hos	AM	0.22	0.32	0.04	0.29	0.60	0.51	0.42	0.44	0.82	**0.82**	0.46	0.74
	PM	0.21	0.28	0.14	0.23	0.54	0.50	0.65	0.30	0.76	0.78	**0.80**	0.53
	All	0.21	0.29	0.04	0.16	0.57	0.49	0.46	0.36	**0.79**	0.78	0.49	0.52

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 I summarizes the performance of these four models. As can be seen, this table contains three sets of performance comparisons: 1) morning peak hour (“AM”, 7:00 - 10:00 AM); 2) evening peak hour (“PM”, 4:00 - 7:00 PM); and 3) the entire test dataset (“All”). This table has four columns: the first column presents bus stop location information; the second column presents fraction of dwell time predictions with an error between -5 and 0 seconds; and the 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, for the most part, the Bayesian predictive model performs at least as good as or better than the other three models. 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 (see Fig. 6), the Bayesian prediction model significantly outperforms the other models (refer to results for Negley Ave at Centre Ave, Centre Ave at Aiken Ave, and

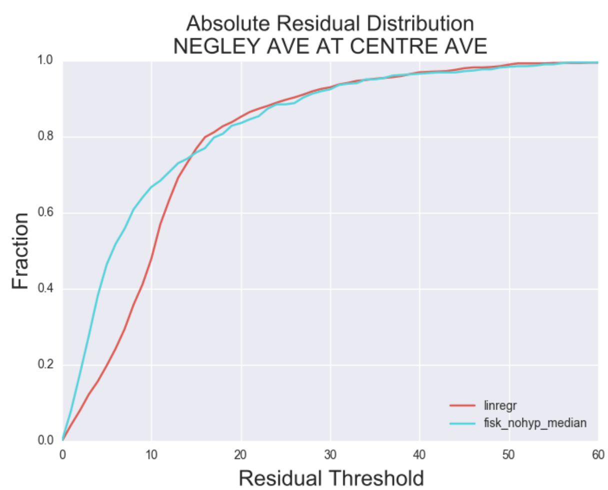

Fig. 7: Fraction of absolute prediction error within a threshold for our framework vs. linear regression. Note that the Bayesian hierarchical model has a higher proportion of small errors.

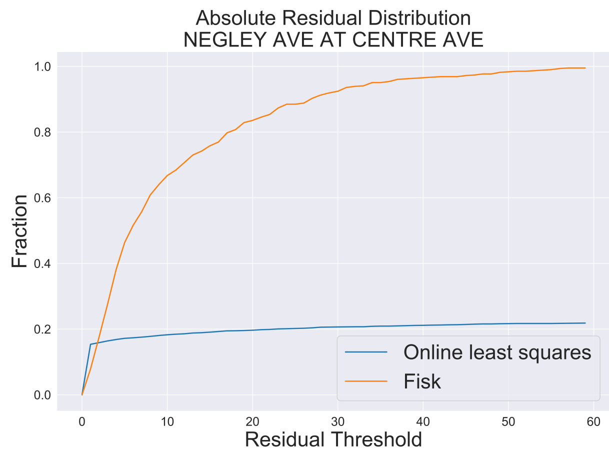

Fig. 8: Fraction of absolute prediction error within a threshold for our framework vs. online least squares regression.

Centre Ave at Craig St NS). Fig. 7 demonstrates this trend for Negley Ave at Centre Ave - the Bayesian model has a much higher proportion of very close predictions than the other error distributions. 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.

C. 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 ($\hat{x}_{\text{on}}$), 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 $\hat{x}_{\text{on}}$. The model details are presented below.

Let Y_i represent the number of passengers boarding 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 $p(\lambda|y)$ is derived as:

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

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

$$\begin{aligned} p(\lambda|y) &\propto p(y|\lambda)p(\lambda) \\ p(\lambda|y) &\propto \lambda^{n\bar{y}} e^{-n\lambda} \lambda^{\alpha-1} e^{-\beta\lambda} \\ p(\lambda|y) &= \lambda^{\alpha+n\bar{y}-1} e^{-(\beta+n)\lambda} \\ p(\lambda|y) &\sim Ga(\alpha + n\bar{y}, \beta + n) \end{aligned}$$

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

A non-informative prior such as Jeffrey's prior is used to bootstrap the system. So $p(\lambda) \propto J(\lambda)^{\frac{1}{2}}$ 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 $\frac{-y}{\lambda^2}$.

$$\begin{aligned} J(\lambda) = -E[\frac{-y}{\lambda^2}|\lambda] &= \frac{1}{\lambda} \\ J(\lambda)^{\frac{1}{2}} &= \frac{1}{\sqrt{\lambda}} \end{aligned}$$

The previous equation can be treated as $Ga(\frac{1}{2}, 0)$. 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 was tested at five out of twelve intersections, and results are summarized in Table II. The results are not bench-marked against any traditional learning model, as the main idea is to demonstrate details of the hierarchical Bayesian framework.

TABLE II: Hierarchical Fisk Model

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

VI. Broader Applicability

Although our principal research interest is effectively utilizing V2I communication of real-time information from buses to improve real-time traffic control decisions, we believe that the Bayesian hierarchical framework presented in this paper has broader applicability to other planning and scheduling under uncertainty problems. To cope with uncertainty in task durations and outcomes, a range of techniques for building resilient plans and schedules have emerged over the years. Some techniques have relied on knowledge of uncertainty limits to generate plans that retain temporal flexibility [31–34].

Others have exploited probabilistic models of task duration and outcome uncertainty to generate plans or policies that optimize expected behavior [35–39]. Still other techniques have used probability distributions to predict durations within deterministic optimization procedures [40, 41]. In all cases however, the effectiveness of these techniques depends on the availability of good probabilistic task models.

There are four primary advantages of using the Bayesian hierarchical framework introduced above. First, it offers robust predictions in highly stochastic and noisy environments, which often have a large variance and noise in both the independent and dependent variables that are incorporated. Second, the Bayesian approach effectively addresses uncertainty by delivering a confidence in the prediction in the form of a posterior predictive distribution. Planning and scheduling systems can then use this confidence to inform their decisions. Third, the framework requires little data, both in the selection and prediction stages. The selection stage involves choosing the likelihood for the task duration 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. 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.

VII. 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 bench-marked against those obtained from both linear and online regression models. The results demonstrate that the Bayesian model is able to perform at least as good as, and in most instances far better than both traditional learning models and recently popular deep 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 $\hat{x}_{\text{on}}$. 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] 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.

[2] S. Sen and K. L. Head, "Controlled optimization of phases at an intersection," *Transportation science*, vol. 31, no. 1, pp. 5–17, 1997.

[3] 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.

[4] M. Eichler and C. F. Daganzo, "Bus lanes with intermittent priority: Strategy formulae and an evaluation," *Transportation Research Part B: Methodological*, vol. 40, no. 9, pp. 731–744, 2006.

[5] G. Zhou and A. Gan, "Performance of transit signal priority with queue jumper lanes," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1925, pp. 265–271, 2005.

[6] G. Zhou, A. Gan, and X. Zhu, "Determination of optimal detector location for transit signal priority with queue jumper lanes," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1978, pp. 123–129, 2006.

[7] H. Liu, A. Skabardonis, W.-b. Zhang, and M. Li, "Optimal detector location for bus signal priority," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1867, pp. 144–150, 2004.

[8] J. Ding, M. Yang, W. Wang, C. Xu, and Y. Bao, "Strategy for multiobjective transit signal priority with prediction of bus dwell time at stops," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2488, pp. 10–19, 2015.

[9] I. K. Isukapati, H. Rudová, G. J. Barlow, and S. F. Smith, "Analysis of trends in data on transit bus dwell times," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2619, pp. 64–74, 2017.

[10] K. Zografos and H. Levinson, "Passenger service time in a no fare bus system," *Transportation Research Record*, vol. 1051, pp. 42–48, 1986.

[11] R. Rajbhandari, S. I. Chien, and J. R. Daniel, "Estimation of bus dwell times with automatic passenger counter information," *Transportation Research Record*, vol. 1841, no. 1, pp. 120–127, 2003.

[12] K. J. Dueker, T. J. Kimpel, J. G. Strathman, and S. Callas, "Determinants of bus dwell time," *Journal of Public Transportation*, vol. 7, no. 1, p. 2, 2004.

[13] A. Tirachini, "Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers," *Transportmetrica A: Transport Science*, vol. 9, no. 1, pp. 28–49, 2013.

[14] F. Martínez, M. P. Frías, M. D. Pérez, and A. J. Rivera, "A methodology for applying k-nearest neighbor to time series forecasting," *Artificial Intelligence Review*, vol. 52, no. 3, 2019.

[15] F. Martínez, M. P. Frías, M. D. Pérez-Godoy, and A. J. Rivera, "Dealing with seasonality by narrowing the training set in time series forecasting with knn," *Expert Systems with Applications*, vol. 103, pp. 38–48, 2018.

[16] H. Tyralis and G. Papacharalampous, "Variable selection in time series forecasting using random forests," *Algorithms*, vol. 10, no. 4, p. 114, 2017.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] F. A. Gers, D. Eck, and J. Schmidhuber, "Applying lstm to time series predictable through time-window approaches," in *Neural Nets WIRN Vietri-01*. Springer, 2002, pp. 193–200.

[19] R. Fu, Z. Zhang, and L. Li, "Using lstm and gru neural network methods for traffic flow prediction," in *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2016, pp. 324–328.

[20] N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, "Time-series extreme event forecasting with neural networks at uber," in *International Conference on Machine Learning*, vol. 34, 2017, pp. 1–5.

[21] K. Bandara, C. Bergmeir, and S. Smyl, "Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach," *Expert Systems with Applications*, vol. 140, p. 112896, 2020.

[22] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Applied Soft Computing*, vol. 54, pp. 246–255, 2017.

[23] M. R. Oliveira and L. Torgo, "Ensembles for time series forecasting," 2014.

[24] 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.

[25] R. Wilcox, "Kolmogorov–smirnov test," *Encyclopedia of biostatistics*, 2005.

[26] 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.

[27] 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.

[28] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The american statistician*, vol. 49, no. 4, pp. 327–335, 1995.

[29] J. Skilling *et al.*, "Nested sampling for general bayesian computation," *Bayesian analysis*, vol. 1, no. 4, pp. 833–859, 2006.

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

[35] M. Drummond, J. Bresina, and K. Swanson, "Just in-case scheduling," in *Proceedings AAAI-94*, 1994.

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

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

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

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

[40] 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.

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

Isaac K Isukapati, PhD is a project scientist at the Robotics Institute at Carnegie Mellon University. Dr. Isukapati is interested in the development of scalable mathematical and statistical models for reasoning under uncertainty. His other research interests include computational game theory, distributed & multi-agent systems, travel time reliability monitoring systems, adaptive signal control.

Conor Igoe completed his bachelors in Electronic Engineering at University College Dublin. Mr. Igoe will start his PhD in the Machine Learning Department at Carnegie Mellon University.

Eli Bronstein is a senior in the department of Electrical Engineering & Computer Science at University of California at Berkeley.

Viraj Parimi is a Masters student in the Robotics Institute at Carnegie Mellon University.

Stephen F. Smith, PhD is a research professor at the Robotics Institute at Carnegie Mellon University. Dr. Smith has over 35 years of experience in technology development for complex planning, scheduling, and optimization problems. Dr. Smith is a Fellow of the Association for the Advancement of AI (AAAI).