

# Neural Networks for Predicting Bus Arrival Times

Cooper Sloan

Massachusetts Institute of Technology  
csloan@mit.edu

Jiahao Chen

Massachusetts Institute of Technology  
jiahao@mit.edu

Alan Edelman

Massachusetts Institute of Technology  
edelman@mit.edu

## ABSTRACT

Current bus networks do not provide accurate schedules, making them unreliable and inconvenient for passengers. Providing passengers with accurate predictions of bus arrival times is necessary to reduce wait times. Neural nets have achieved high accuracy with root mean square error (RMSE) of around 60 seconds in predicting bus arrival times when compared with historical and statistical models. However previous research has used small datasets which span only a few months. This only allows for relatively small models with a few hidden layers and tens of nodes and prevents studying more complicated effects. This research uses a larger neural net trained by a dataset which spans several years, in order to improve accuracy of predictions and study periodic effects. We find that periodic effects are dominated by intrinsic properties of the bus network, and that larger neural networks achieve similar accuracy to current research, with RMSE from 50-80 seconds.

## INTRODUCTION

Bus schedules need better reliability for passengers in order to increase the appeal of public transport. Most workers commute alone in private cars, and only 10.8% use public transport. Commute times for US workers have increased steadily over the past several decades despite increases in transit system capacity [1]. Public transportation can reduce these commute times by improving network efficiency. Public transportation also has many other benefits including reduced pollution, traffic, and cost for commuters. Nevertheless, buses are less convenient than private cars because they run behind schedule and are unreliable for commuters. Providing accurate arrival time predictions to passengers decreases wait times for passengers, and increases the overall efficiency of the bus network.

The problem of predicting bus arrival times is as follows. Passengers riding the bus would like to know when the bus will arrive at each stop so that they can

catch the bus on time and minimize the time they have to wait. Buses publish schedules, but they are typically only available for a subset of stops, sometimes called checkpoint stops. When a bus reaches a checkpoint stop, it will wait if it is ahead of schedule, so that it can remain consistent with passenger expectations. Nevertheless, buses rarely stay on schedule, and schedule deviations tend to propagate. In practice bus schedules are rarely reliable.

It is therefore very valuable for passengers to have accurate predictions of bus arrival times. These predictions can be derived from several known factors in order to predict future arrival times. The factors used to predict include arrival times at previous stops, current and average velocity, travel times between stops, passenger counts, and current network load. Only a subset of these factors may be available depending on the bus network being studied. The most important predictions are the times a bus will arrive at then next several stops. Prediction accuracy tends to go down for stops very far in the future; however passengers typically will only desire to know the next few upcoming arrival times. Another consideration for generating predictions for bus arrival times is that algorithms need to be reasonably fast in order to provide utility for passengers. Predictions should be delivered to passengers in real time so that they have knowledge of when to arrive at the stop. This information is typically delivered via mobile apps such as Google Maps, and not typically available from the bus network authority themselves. Predictions should also be able to be updated in an online fashion as conditions change. Traffic can be very volatile and predictions need to be updated quickly. This limits the types of models which can be used.

Many techniques have been used to try to predict arrival times to reduce waiting times for passengers, including historical models, regression, and neural networks [4]. A few key features of bus networks suggest that sophisticated models may be effective.

Buses traffic is more difficult to predict than normal traffic because buses have to adhere to a schedule and pick up passengers. Additionally, buses in a route exhibit a phenomenon called clumping. As a bus starts to run behind schedule, more passengers will arrive at future stops. The bus will then have to spend more time at future stops to pick up passengers, which causes it to fall more behind schedule. The next several buses will then catch up to that bus, causing the buses on that route to clump together. These complicated effects cannot be captured by simple models. Sophisticated statistical models and machine learning models like neural networks have shown the best success in predicting arrival times [4]. The state of the art models have an RMSE around 60 seconds when predicting a few stops ahead. However, current research works with relatively small data sets. This prevents incorporating periodic effects into the model.

This research uses neural networks to improve upon prediction accuracy by exploiting periodic effects. Time series data like bus arrival times exhibits periodic behavior across days, weeks, months, and years. For example, traffic surges during rush hour in a predictable pattern throughout the day, and holiday transit patterns repeat every year. Our model is trained using Massachusetts Bay Transportation Authority (MBTA) bus GPS data which spans several years. This is the largest dataset used for predicting bus arrival times. Models are trained across periods to determine whether neural nets can pick up on periodic effects to improve the prediction of arrival times. We find that our models were not able to significantly pick up on periodic effects, however they achieve similar accuracy to the state of the art.

## PREVIOUS WORK

Neural networks have shown the best accuracy for predicting bus arrival times [5]. Studies have been conducted in several locations throughout the world including several cities in the US as well Canada [2], China [1] and Brazil [4]. Several techniques have been used to predict bus arrival times including historical models, statistical models, and machine learning methods. Even with temporally limited data, neural networks are very effective at predicting arrival times [3]. A study in Texas compared the accuracy of neural networks as compared with historical and

regression based models and found that neural nets produce around twice the accuracy compared to the other models [5]. For this reason our study uses neural networks.

Previous studies using neural networks use a few hidden layers with small amounts of data. Chien, Ding and Wei used artificial neural networks (ANNs) to predict bus arrival times [3]. However the network architecture was the most basic possible, with a single input layer, hidden layer and output layer. Even so, their models show good accuracy and this research borrows the same feature vectors. More sophisticated neural nets have been shown to be effective in other use cases in recent years [6]; however they have not been applied to this problem. The complicated patterns and phenomena in bus networks suggest that large models may be effective. Because of this we use larger neural nets with more nodes and hidden layers.

Current research to predict bus arrival times uses limited data sets which span around a month and are not suitable for analyzing long term seasonal effects. Large bus GPS data sets are rare, so some studies get around this by simulating data [3]. A study in China uses data spanning a month [1]. Deep neural nets particularly require a large amount of data, and the lack of data motivates studies to use smaller models to avoid overfitting. However, the size of the MBTA dataset allows for larger models which may perform better compared to previous models.

Our research borrows the feature vector set from previous research and uses neural nets to model arrival times. However, it departs from previous research by utilizing a larger dataset as well as larger models in order to exploit seasonal effects and improve the accuracy of predictions.

## FEATURE SELECTION AND EXTRACTION

This research uses publicly available GPS data from the MBTA to calculate arrival times for buses at each stop. The data set consists of three main sections. First it contains information about bus routes with the name, latitude and longitude of stops. Second it contains information about bus schedules for weekdays, weekends and holidays for each route. Finally it contains GPS data for every bus on the network, along with metadata about which route the

bus is on and whether it is inbound or outbound. The GPS data is obtained by polling the MBTA server periodically. The amount of GPS data available spans several years. However schedules for bus routes change often and archived records are not publicly available from the MBTA. Additionally, bus stop locations change over the years, and it is difficult to determine the location of a bus stop a past point in time. Furthermore, GPS data can be inaccurate especially in urban valleys which are common in an urban area like Boston. All of these considerations were taken into account during feature extraction.

Precautions were taken while extracting the features from the MBTA GPS data. First, data from only the previous two years were used to minimize the amount of stops changing location as well as schedule changes. Even with this partitioning, the timescale of data is much longer than that of previous research. The route selected for testing is relatively long, with 27 stops, and it very straight, with only a few turns at the beginning and end. Minimizing the number of turns ensure that GPS data does not overlap. For example, if a bus has to make a U-turn, it can be difficult to tell which direction it was heading from the raw GPS data. For this reason we only consider the stops in the middle of the route when training the model.

To determine the arrival time at each stop, GPS locations of the buses are compared to the GPS location from the bus stop data set. After the bus is within a certain threshold of the stop, the time is noted. Due to inaccuracy of GPS data, this threshold must be tuned. A higher threshold will capture more of the data but arrival times will be less accurate. A lower threshold will yield more accurate arrival times, however some of the buses will never read GPS locations within the threshold due to noise, leading to a missing value for its arrival time. The inputs to the neural net consider one full trip of a bus from its first stop to its final stop. Therefore only buses which have a recorded arrival time at every stop are considered as input data to the model. About half of the trips met this criteria, which was plenty for testing and evaluating the model.

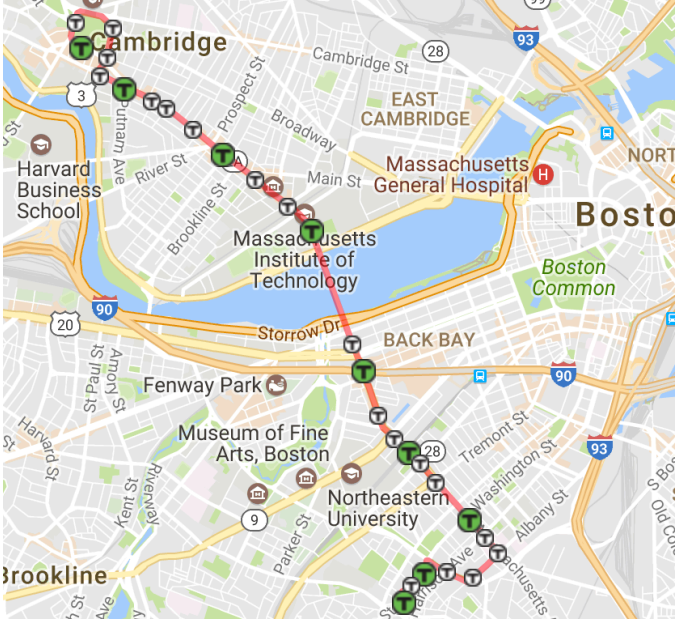
The bus data were partitioned in order to capture trends in bus traffic. Because buses have different schedules on weekdays and the weekend the data

were partitioned according to weekday, Saturday or Sunday. Buses arrive at different intervals throughout the day, so the data were partitioned so that in each partition the buses at a given stop arrive at the same intervals. Finally, the data were partitioned by month and year so that seasonal effects could be studied.

The features selected for the neural net were the travel times for a bus between each stop along the route. Several features are commonly used to predict bus arrival times. Travel time, or the difference in time between each stop on a route is almost always used. Arrival time, or the time it takes to get from the first stop in the route to each successive stop is also common. Additionally neural nets use dwell time, or the amount of time a bus stays at each stop before moving on. Finally some models use schedule adherence, or the difference between the arrival time at each stop and the scheduled arrival time by the bus system. Due to the difficulty of determining bus schedules across years, schedule adherence was not used as a feature as it would be unreliable. Arrival time does not work well with neural nets because for a long route, the magnitudes across stops will be very different, weighting later stops much more. Travel times are much more similar in magnitude, so travel times were used instead of arrival time. Dwell time is difficult to compute from GPS data so it was not used.

## CASE STUDY

In order to test the model, the Route 1 bus was selected. It runs from Harvard in Cambridge to Dudley Station in Boston via Massachusetts Avenue, passing 25 stops along the way. The stops vary in distance, with the longest around half a mile and the shortest several hundred meters. This route was chosen because it is well traveled and relatively long compared with other MBTA routes. Only inbound buses were analyzed.



**Figure 1:** 1 bus route in Boston from Harvard Square to Dudley Station

## MODEL DEVELOPMENT

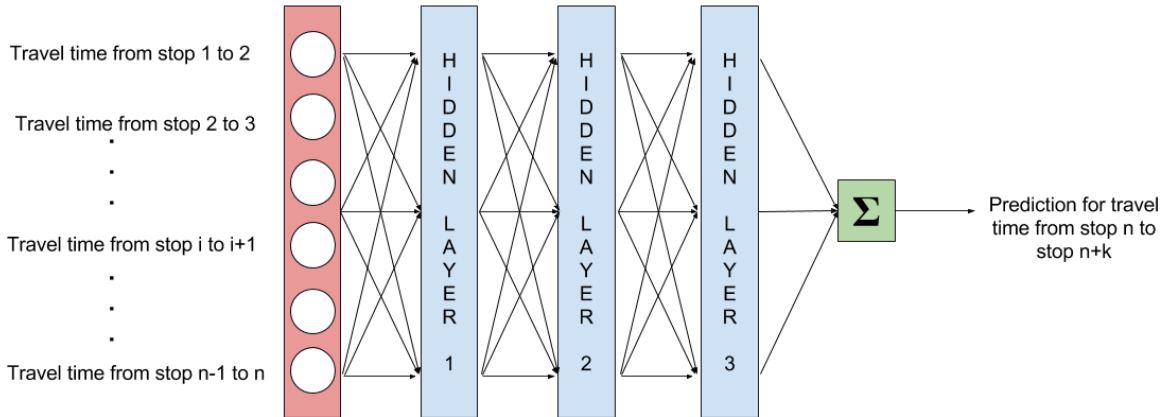
Several possible model architectures have been suggested for predicting bus arrival times. For an overview of modeling for bus networks see [3]. The final model architecture used three fully connected hidden layers with Sigmoid activation functions. This architecture is visualized in Figure 2. The inputs to the neural net were the travel times between stops, and the output is the predicted travel time to the next stop. By simply adding the prediction to the arrival time at the last stop, one can calculate the predicted

arrival time at the next stop. Furthermore, the network can predict the travel times to stops even further down the route.

The actual value of these travel times can be computed as follows.  $A_j$  is the arrival time at stop  $j$ ,  $T_{ij}$  is the travel time from stop  $i$  to stop  $j$ , and  $P_{ij}$  is the predicted travel time from stop  $i$  to stop  $j$  as computed by the neural net. If the net is predicting the arrival time  $m$  stops down the line, the arrival time can be computed as the offset from the beginning of the route plus the sum of travel times between all of the known stops, plus the predicted travel time to stop  $j$ .

$$A_j = A_0 + \left( \sum_{i=1}^{j-m-1} T_{i,i+1} \right) + P_{j-m,j}$$

One would expect the accuracy of these predictions to decrease as the predictions are further in the future. This is because as you travel further down the route, random variations in the traffic conditions accumulate, leading to higher variance in the arrival times. Due to this reason, most systems can only provide accurate predictions for around 3-8 stops before the error reaches a level which is akin to random guessing. So two hyperparameters of the model are the number of input travel times, and the number of stops ahead to predict. Including too few input travel times does not give the neural net enough information to accurately predict. However, travel times for stops which are much before the stop to predict will have little impact on the final travel time,



**Figure 2:** Three layered neural network with fully connected layers and Sigmoid activation functions

and thus will essentially be noise for the neural net. After some tuning, the number of input travel times to use was set to 10.

Sigmoid activations functions are commonly used for transportation systems so they were selected [5]. Most current research used a single fully connected hidden layer, however they use a limited amount of data so using more layers would subject them to overfitting. This research used multiple hidden layers with in order to attempt to capture the intricate periodic patterns in the bus data. Because the neural net produces real valued outputs, a square loss was used to generate predictions. This loss function is obtained as follows.

$$\frac{1}{2N} \sum_{i=1}^n |y_i - p_i|^2$$

Where  $p_i$  is the predicted travel time and  $y_i$  is the actual arrival time for bus  $i$ .

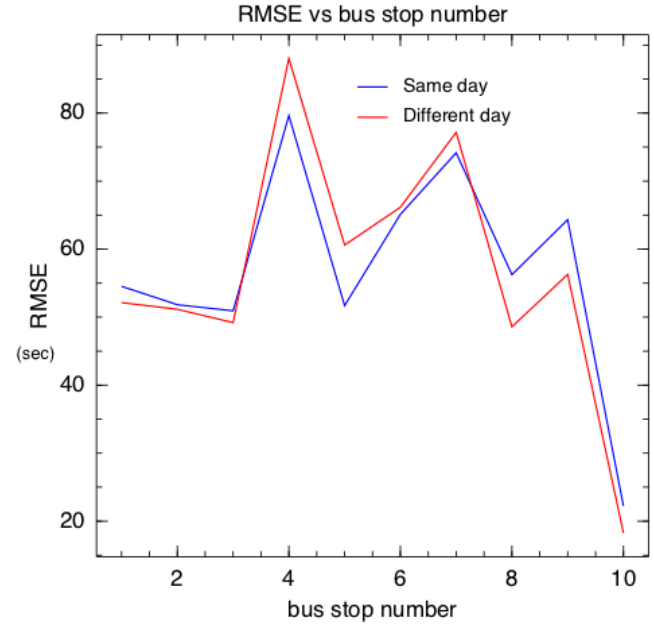
Standard stochastic gradient descent was used to train the model, with a fixed learning rate of 0.0001, a mixed momentum of 0.9, and a regularization coefficient of 0.0005.

## MODEL EVALUATION

In order to test whether neural net can utilize periodic effects to improve bus arrival time prediction accuracy, the training and validation data were partitioned across the desired period. In this case the periods being studied were across months and weeks. First, a baseline accuracy was established using a large training set of around 5000 data points which span around a year. This is a larger training set than most previous studies, which use several hundred data points.

The first test run assessed whether neural nets could pick up on patterns throughout the week. In order to determine this, the training set was selected from a series of data from Mondays. Two test sets were then selected. The first set was from the Monday from a week later. The second set was from the Tuesday from a week later. If the neural net was able to pick up on periodic effects, one would expect the

validation accuracy from the first set to be higher than the validation accuracy from the second set. The accuracy can also be compared across different days of the week in order to assess the correlation between different days of the week. If two days are highly correlated, data from either of them can be used to predict the other. The accuracy of each set is shown in Figure 3.

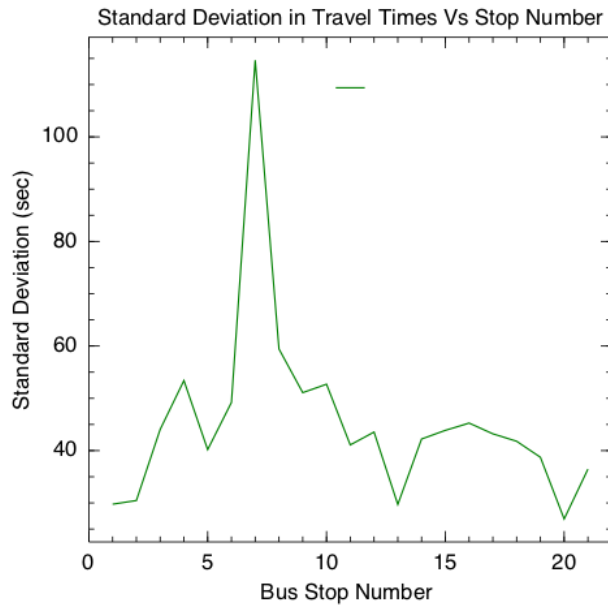


**Figure 3:** Validation Root Mean Square Error for buses on the same day and different days from training data

The first observation is that large neural nets are effective at predicting bus arrival times. The root mean square error in the graph corresponds to providing predictions which are accurate to plus or minus a minute. The average RMSE across the stops is 63.6 s, which is comparable to 61.43 s figure from the Chien study [3]. The second observation is that periodic effects are not significantly captured by this model. If the neural net was able to pick up on weekly patterns, the blue curve should fall below the red curve. This is not the case. Instead both curves follow roughly the same path, indicating that test prediction quality is somewhat independent of which day training data comes from.

Although the neural net does not pick up on periodicity, the similarity between the two curves

suggests something about the nature of predicting arrival times. In order to understand this it is necessary to look at the variance in bus travel times between stops. The standard deviation in travel times between bus stops is shown in Figure 4.



**Figure 4:** Standard Deviation in travel times between stops

The standard deviation varies noticeably between different links in the network. In particular, the standard deviation at checkpoint stops is very low. This can be seen in the graph at for example stop 1, 13, and 20. Additionally, a large peak can be seen at stop 7. This link in the route happens when the bus crosses the Harvard Bridge across the Charles River, which is a bottleneck for traffic entering Boston from Cambridge. This explains the high variance. Finally, comparing the variance of travel times between stops with the prediction error, it is clear that these factors dominate any periodic effects which may be present. Additionally, because weekly effects tend to be stronger than monthly and yearly effects, neural nets may not be suitable for analyzing more long term effects. High variance links lead to high error in estimation due to more stochastic effects occurring at those portions of the network. Despite comparing different days of the week in the validation sets, both sets achieve roughly the same accuracy. This indicates that seasonal effects are dominated by

intrinsic properties of the bus route. The properties which lead to different variances amongst links in the route are number of stoplights, number of lanes, and outside traffic.

The similarity between the RMSE from this study and other studies using smaller nets may suggest an upper bound on the accuracy which can be obtained in bus arrival time predictions. Neural nets obtain higher training accuracy when given more hidden layers [5], and typically perform better with a larger amount of training data. The lack of an accuracy improvement in this research compared with previous research suggests either that the neural network architecture is not properly tuned, or prediction accuracy is limited. It is possible that an improved neural net may yield higher accuracy. For example, the feature vectors used in this research was limited, using only travel times. More sophisticated features such as dwell time and schedule adherence have shown good results [3], however it was not reasonable to calculate these features given the accuracy of the GPS data. Despite not using a limited feature set, this research obtains similar error measures to previous research.

Another important observation is that neural nets yield much better results when compared to regression models and historical data based models [5]. This suggests that neural networks are the right approach for this problem, and that the results produced by neural nets are pushing the limits of what is possible. However there are certain features of a traffic network that neural nets cannot predict. For example they cannot predict the occurrence of an accident which would dramatically slow down traffic. Although this is an extreme situation, there are similar stochastic effects which neural networks cannot predict, such as the timing of stoplights. A single stoplight can easily slow down a bus by 30 seconds within a single route. In the case of the 1 bus route, a bus passes through as many as 6 stoplights between two bus stops. This contributes to the high standard deviations in travel times as shown in figure 4.

Another key factor that affects arrival is passenger count. Buses have to wait for passengers to get on and off, but buses do not necessarily stop at every stop if no one gets on or off. Predicting which stops a bus will stop at and how long it will stay at each stop requires knowledge of the passengers on the bus,

rather than just knowledge of the trajectory of the bus. This suggests that GPS based data sets may not be sufficient to more accurately predict arrival times.

The final consideration is whether more accuracy is desired. Bus arrival time predictions are useful because they can be distributed to passengers via a mobile app such as Google maps so that they can plan to arrive at the stop on time. Research has shown that neural nets can consistently predict arrival times within about a minute [3], and this level of accuracy may be accurate enough more most passengers.

## CONCLUSION

To provide accurate predictions of bus arrival times research uses a neural network with three hidden layers trained on data from the Route 1 bus in Boston. The scale of the dataset which trains the network is around 10 times larger than previous studies, which use a few months of data rather than years. Any periodic effects that may be present in the long term data are dominated by local effects of the network such as variance in travel time between stops. The error rates observed are on par with those of previous research. We find that large models trained on long term data can accurately predict bus arrival times with RMSEs of around 60 seconds. These predictions can be delivered to passengers to reduce waiting times and improve network efficiency.

## REFERENCES

- [1] McKenzie, Brian, and Melanie Rapino. Commuting in the united states: 2009. US Department of Commerce, Economics and Statistics Administration, US Census Bureau, 2011.
- [2] Patnaik, Jayakrishna, et al. 2004. Estimation of Bus Arrival Times Using APC Data. *Journal of Public Transportation*, 7 (1): 1-20.
- [3] Chien, S., Ding, Y., and Wei, C. (2002). "Dynamic Bus Arrival Time Prediction with Artificial Neural Networks." *J. Transp. Eng.*, 10.1061/(ASCE)0733-947X(2002)128:5(429), 429-438.
- [4] Jeong, Ranhee, and Laurence R. Rilett. "Bus arrival time prediction using artificial neural network model." *Intelligent Transportation Systems*, 2004. Proceedings. The 7<sup>th</sup> International IEEE Conference on. IEEE, 2004.
- [5] Hagan, M. T., Demuth, H. B., and Beale, M. ~1996. *Neural network design*, PWS, Boston
- [6] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *Nature* 521.7553 (2015): 436-444.