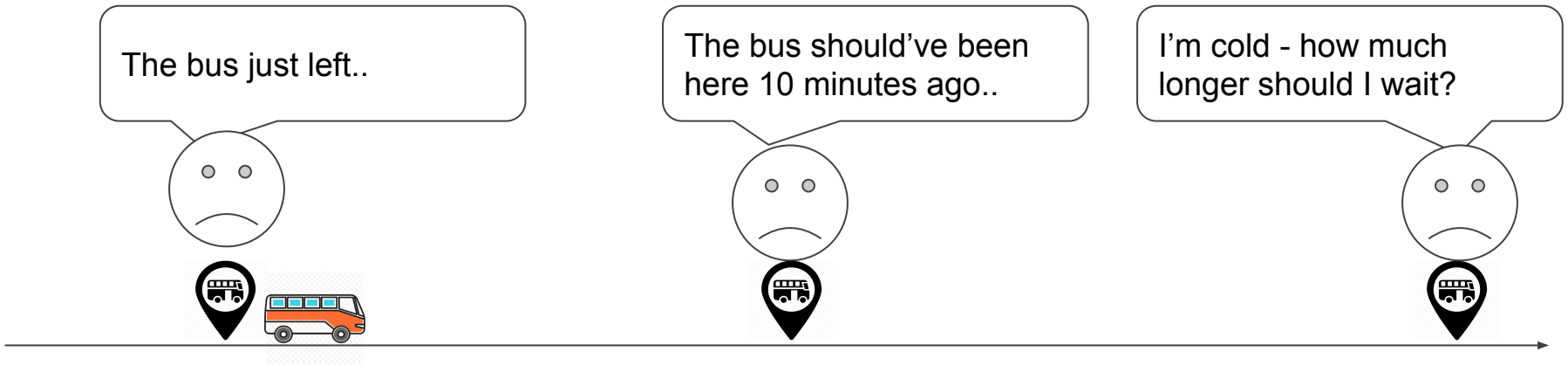
Arrival Time Prediction

Brandon Houghton, Kenji Yonekawa
08-537 - Fei Fang

Background


The bus just left..
The bus should've been here 10 minutes ago..
I'm cold - how much longer should I wait?

Motivation

- High demand for bus arrival prediction
  - + 25,000 installs in Pittsburgh alone
  - Consistent daily useage
- Inefficiency of bus routes
- Lots of available data

Motivation

- High demand for bus arrival prediction
- Inefficiency of bus routes
  - Higher Utilization drives efficiency
  - Real-Time tracking can inform route changes
  - Delay cited as number one deterrent
  - Missing bus due to inaccurate real time info was number three
- Lots of available data

Motivation

- High demand for bus arrival prediction
- Inefficiency of bus routes
- Lots of available data
  - GPS Bus location tracking
  - Passengers track stops within app
  - Real-time traffic estimation
  - Weather, Events, etc.

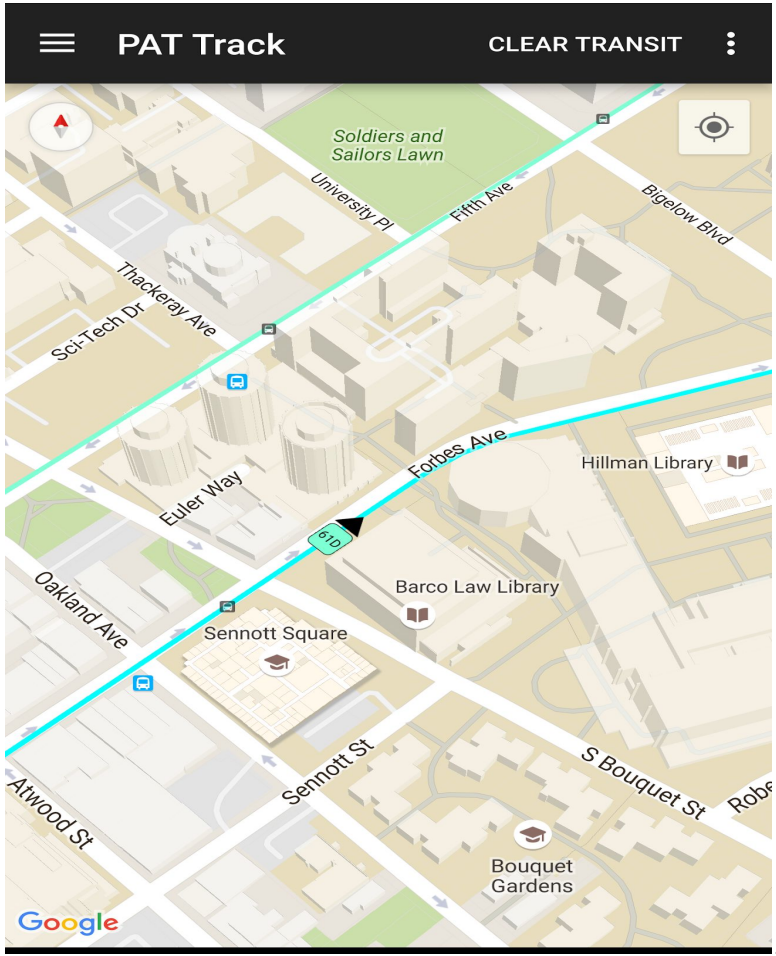
Related Work

- Multiple studies done
  - Y. Bin and Y. Zhongzhen and Y. Baozhen, "Bus Arrival Time Prediction Using Support Vector Machines", 2006
  - W. Treethidtaphat, W. Pattara-Atikom, and S. Khaimook, "Bus Arrival Time Prediction at Any Distance of Bus Route Using Deep Neural Network Model", 2017
  - J. Lei, D. Chen, F. Li, Q. Han, S. Chen, L. Zeng, and M. Chen, "A Bus Arrival Time Prediction Method Based on GPS position and Real-time Traffic Flow", 2017
  - P. Zhou and Y. Zheng and M. Li, "How Long to Wait? Predicting Bus Arrival Time With Mobile Phone Based Participatory Sensing", 2014
- Problems
  - Often use erroneous location tracking
  - Based on a couple of days of data collection
  - Different setting used (location, time, data)

Data Set

- PAT TrueTime API
  - GPS of bus location
  - Updates every 10 seconds
  - Arrival Estimates
- Weather API
  - Precipitation, temperature, wind
  - Updates every hour

Stored in GCP



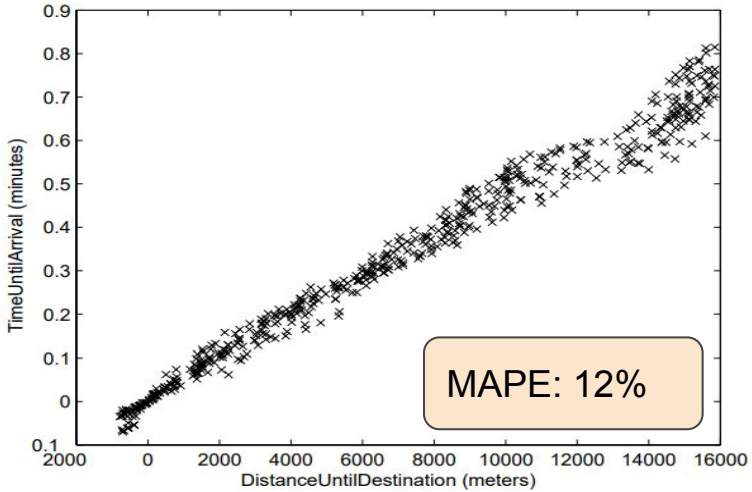
Evaluation

- Limited Horizon
  - Riders not interested in accuracy after 15 minutes
  - Buses can change routes
- Mean Absolute Percentage Error
  - Most common
  - Easy to compare

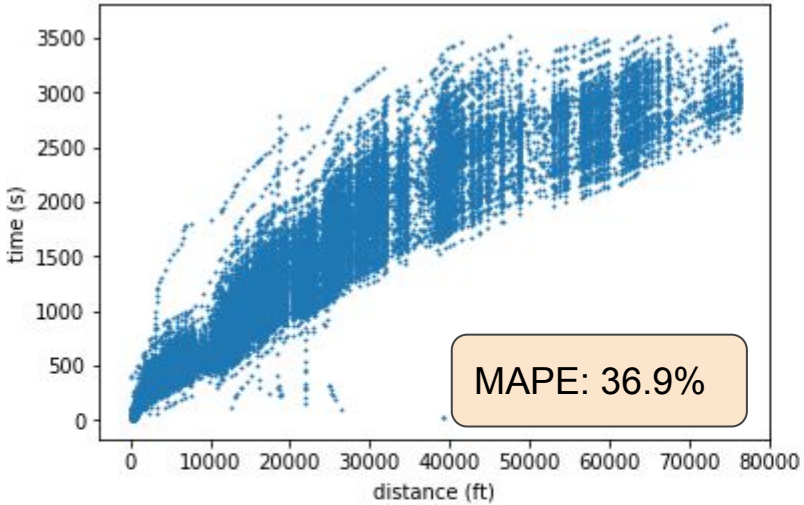
$$\text{MAPE} = \frac{\sum_{t=1}^{n} \left| \left(A_t - F_t \right) / A_t \right|}{n}$$

Evaluation

- Previous approaches don’t generalize well
- Pittsburgh is much more diverse

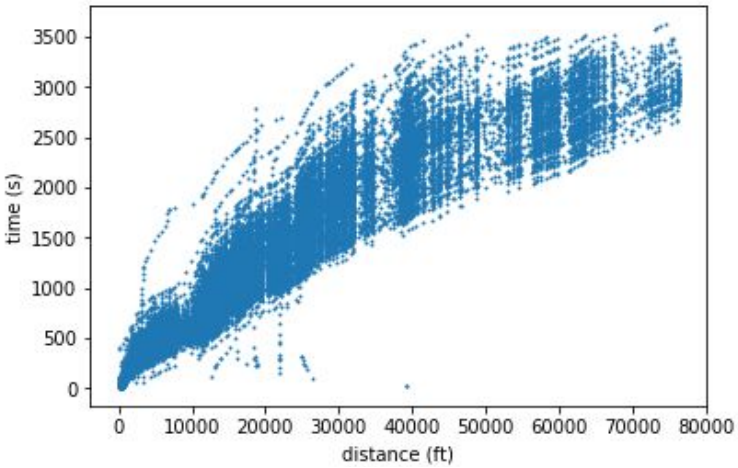


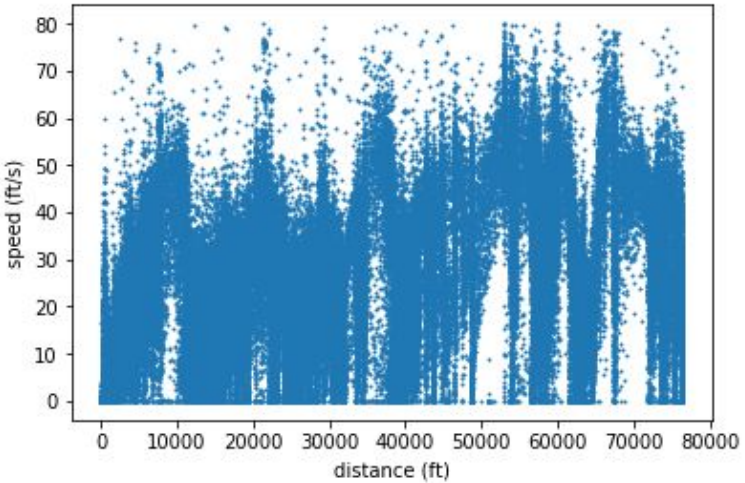
Wall et. al.



Use Velocity Instead of Time

- Errors do not accumulate
- Velocity

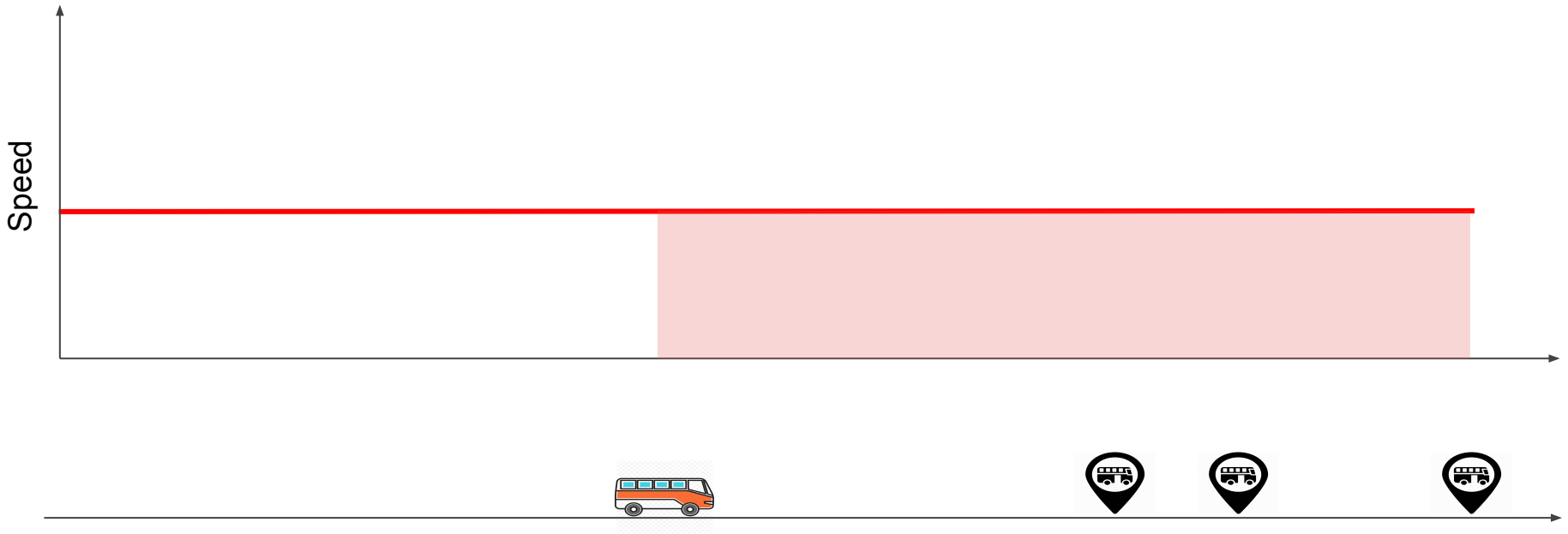




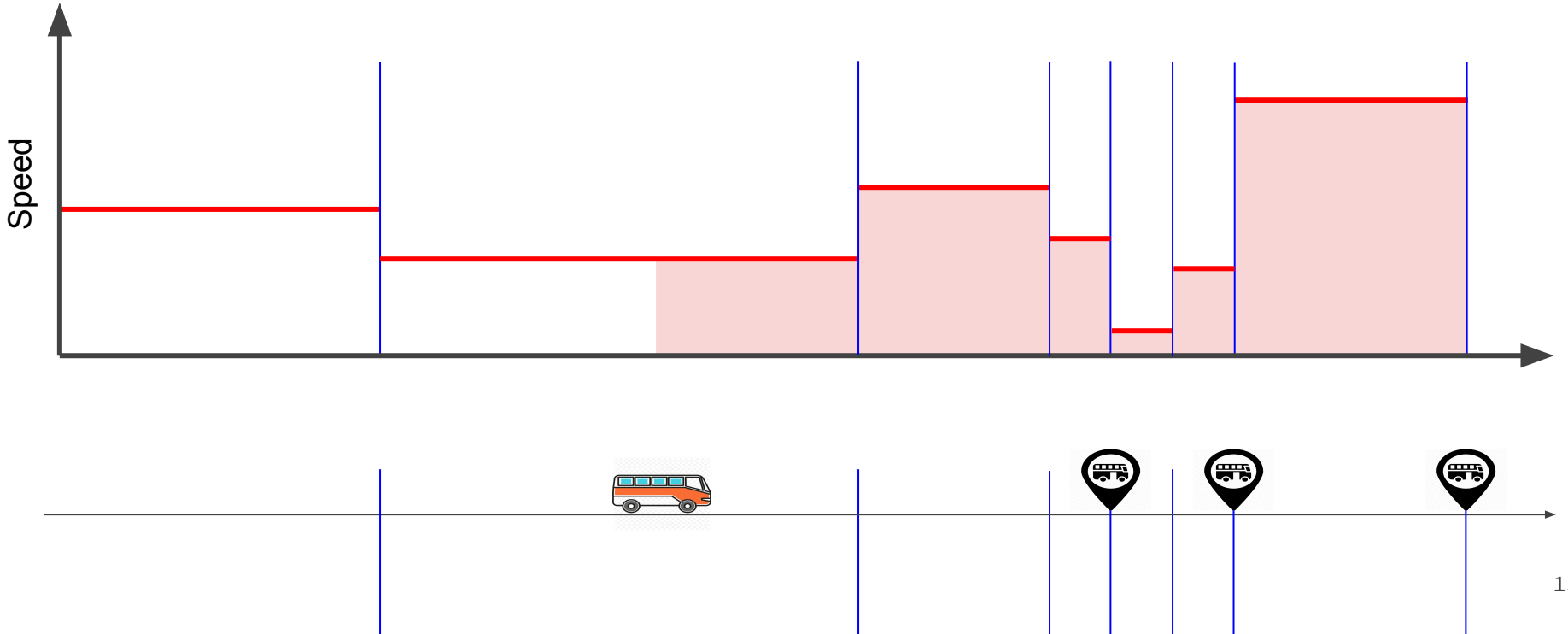
Qualitative Results

- Linear Model
  - Highly dependent on number of bins
- Tree Based Model
  - Does not generalize well to new month of data
- Mixture Models
  - Feature selection was overfitting validation set

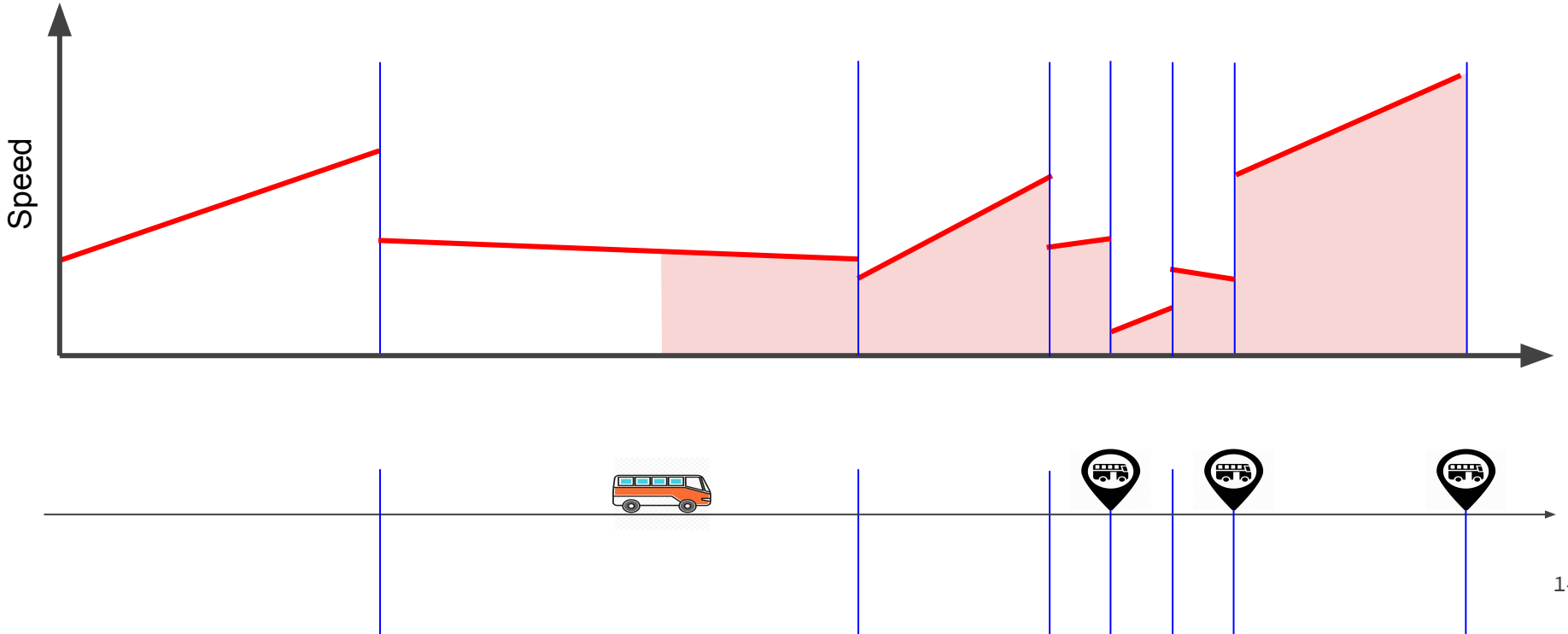
PAT Model



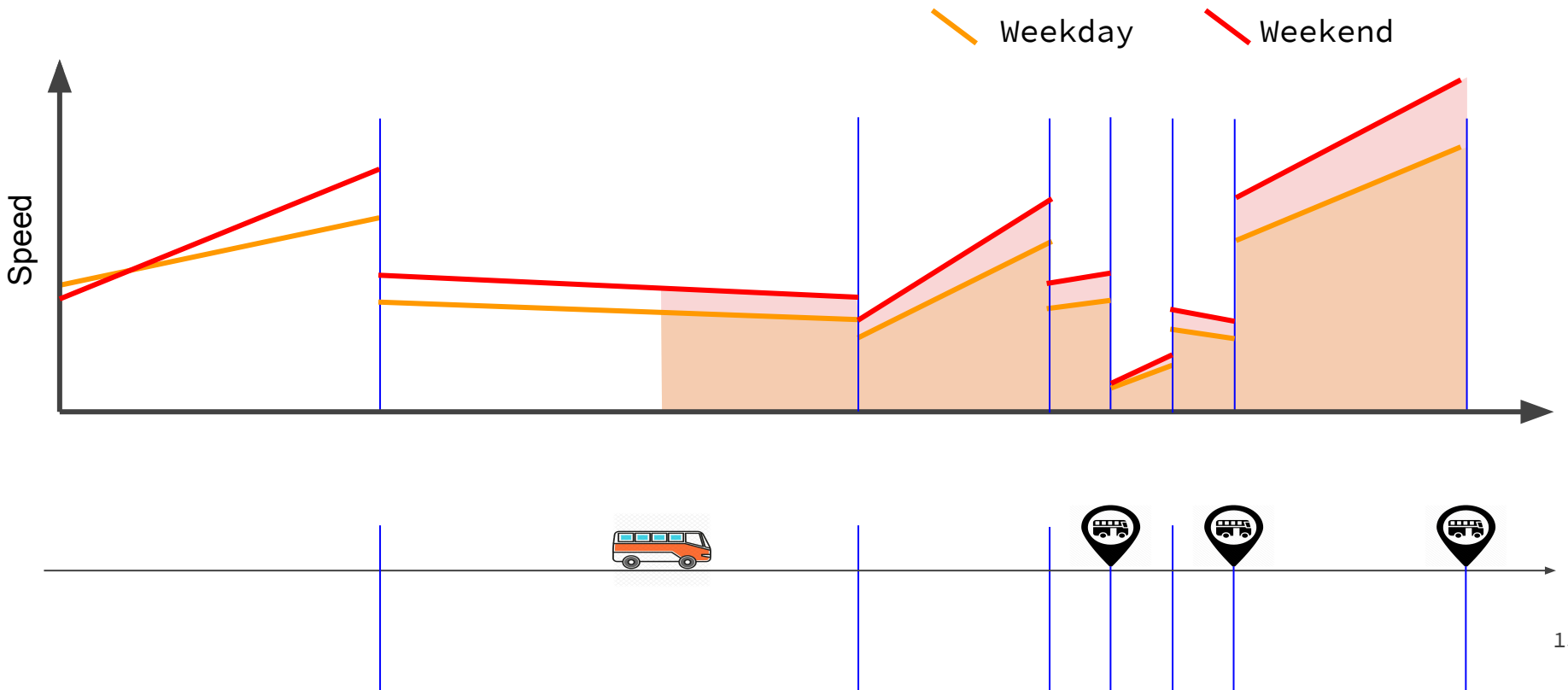
Model Refinement - Tree Segmentation



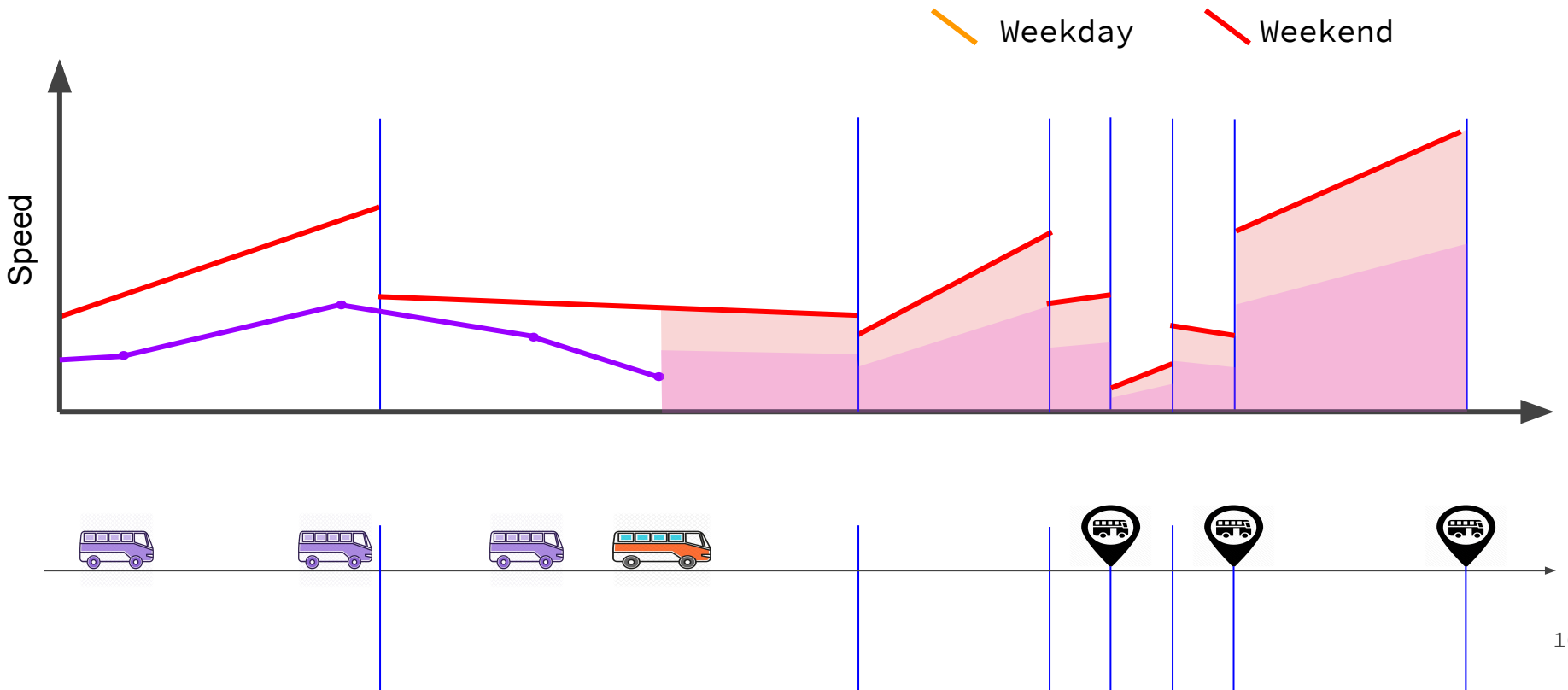
Model Refinement - Linear Interpolation



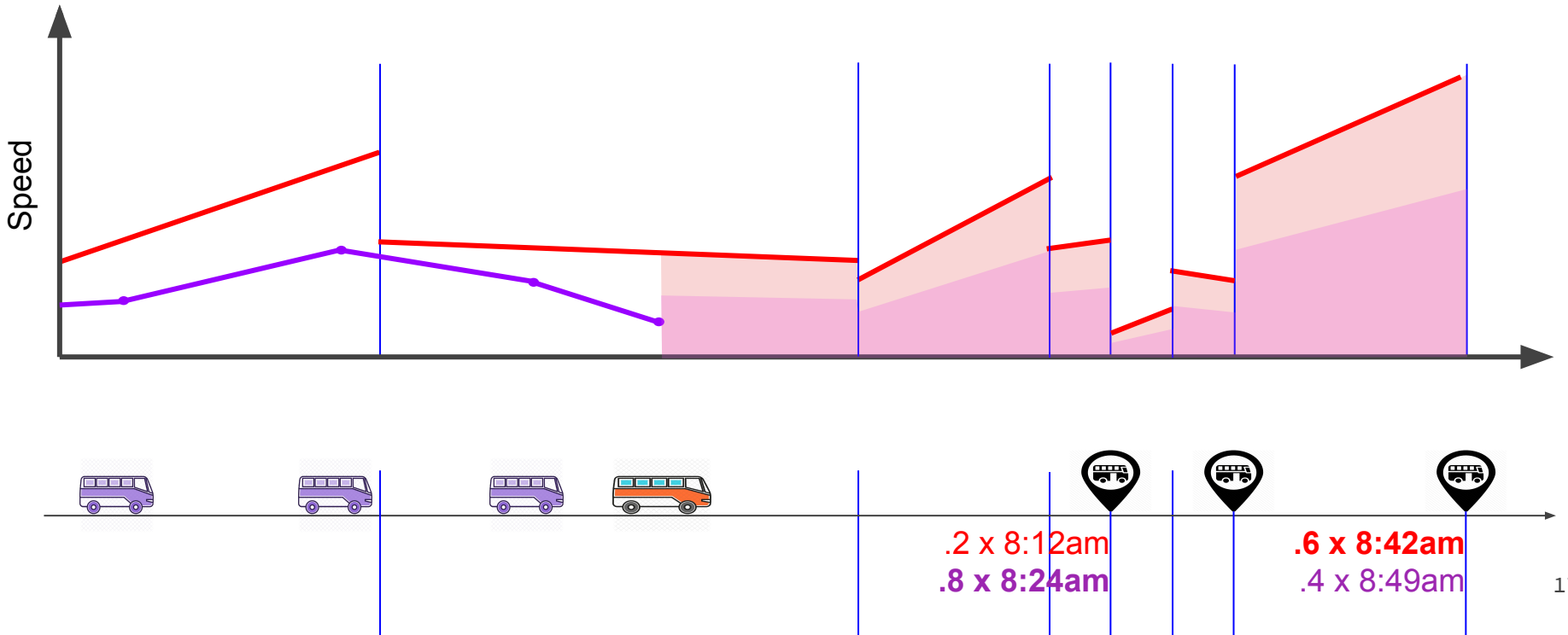
Model Refinement - External Observations



Model Refinement - Momentum Model



Model Refinement - Meta Model



Evaluation Setup

- Results on one Route (61C) that had the most data.
- Train data: March 2018: 155,398 data points
- Test Data: April 2018: 101,504 data points
- True label: future data acquired from PAT’s API

Evaluation Results

#	Model name	Mean Absolute Percentage Error (MAPE)
1	Linear model	28.48%
2	Piecewise linear model	23.84%
3	Decision tree linear model	22.70%
4	Piecewise linear mixture model	18.53%
5	Decision tree with linear mixture model	15.60%
6	Piecewise linear model with momentum	12.25%

PAT's prediction model

Largely affected by historic data. (slope, intercept becomes negative)
Haven't figured out optimal prediction model

Future Work

- Add other data to aid prediction
  - Traffic data from Google’s Real Time Traffic
  - Class schedule for local colleges
  - Holidays and Events
- Provide our contribution as an API or incorporate with smartphone applications
  - Allow applications to integrate improved data without changing apps

Summary

- We tackled the societal challenge of predicting bus arrival time
- Evaluated existing research approaches
- Benchmarked existing API
- Developed and evaluated new approach
- New approach outperforms existing API

Thanks for listening!

Reference

Photo Credits:
https://www.nextpittsburgh.com/city-design/port-authority-rolls-out-real-time-bus-tracking/

Y. Bin and Y. Zhongzhen and Y. Baozhen, "Bus Arrival Time Prediction Using Support Vector Machines", 2006

W. Treethidtaphat, W. Pattara-Atikom, and S. Khaimook, "Bus Arrival Time Prediction at Any Distance of Bus Route Using Deep Neural Network Model", 2017

J. Lei, D. Chen, F. Li, Q. Han, S. Chen, L. Zeng, and M. Chen, "A Bus Arrival Time Prediction Method Based on GPS position and Real-time Traffic Flow", 2017

P. Zhou and Y. Zheng and M. Li, "How Long to Wait? Predicting Bus Arrival Time With Mobile Phone Based Participatory Sensing", 2014

Z. Wall, D. J. Dailey, “An Algorithm for Predicting the Arrival Time of Mass Transit Vehicles Using Automatic Vehicle Location Data”, Transportation Research Board 78th Annual Meeting January 10-14, 1999