Real-Time Bayesian Micro-Analysis Model for Metro Traffic Flow

Eric Lin[1], Jinhyung D. Park[2], Andreas Züfle[3]

[1]Thomas Jefferson High School for Science and Technology [2]The Hill School [3]Geography and Geoinformation Science, George Mason University

Abstract

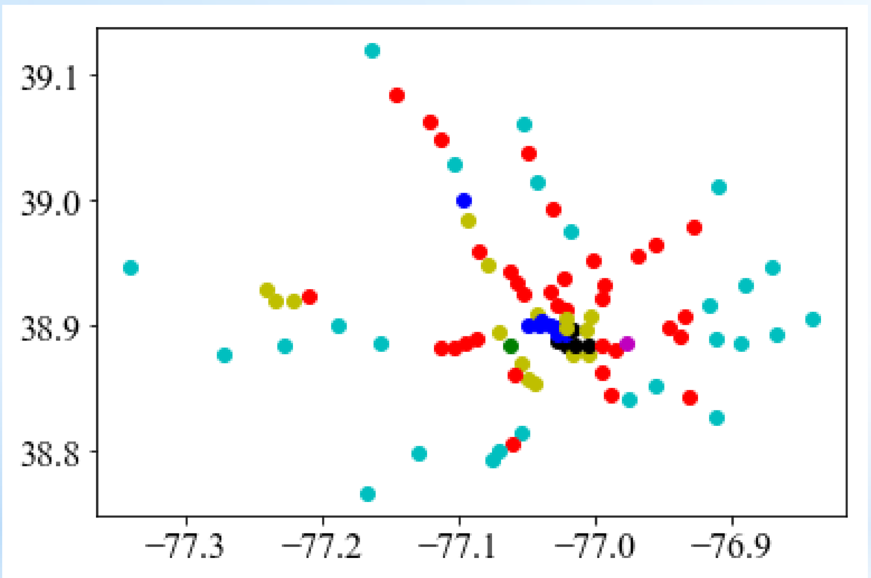
Metro transport plays a large role in major cities around the world as an easily accessible and convenient means of transit. We propose a novel approach to forecast the metro network flow of passengers, which is exceptionally useful for city planning. For instance, the D.C. Metro is currently adding both new trains and stations to its network. Accurate estimations of passenger outflow provide valuable insight in deciding where and when to add new trains and stations.

We present a triple layer micro-prediction approach to predict a passenger's destination time and a station. The first layer develops a probability distribution of origin-destination (OD) stations using analysis on historical data. The second layer conditions the OD probability distribution by destination station and current travel time of a passenger. The conditioning employs spatio-temporal machine learning to improve accuracy. Finally, the third layer conditions the OD probability distribution to the current travel time of a passenger. The culmination of the three layers allows the model to generate the probability of a passenger arriving at any given station and time. For each station, the summation of the probability partition of each passenger in the network produces the expected outflow.

We find that our model outperforms baseline approaches, especially when predicting further into the future. Furthermore, our model can be successfully implemented for a wide array of passenger traffic flow data for smart city planning.

Introduction

- Public transit trip volume is up 34% since 1995, which is higher than the population growth of 21% [2].
- Past research employed machine learning algorithms to model station in and out flow based on large-scale patterns [1].



Definitions

Definition 1: *Trip.* Let $tr = (S_o, S_d, t_o, t_d)$ denote a metro fare card record. S_o, S_d, t_o, t_d denotes the origin station, destination station, time when passenger enters S_o, and time when passenger exits S_d respectively.

Definition 2: *Network Flow Database.* Let $DB_\tau = \{tr_1, \ldots, tr_K\}$ denote a database of in and out flow of a public transit network.

Definition 3: *Meterology.* Let $M_\tau = (p_\tau, h_\tau, d_\tau, v_\tau, r_\tau, e_\tau)$ denote the meteorological features of a time period

Definition 4: *Outflow.* Let $O_{E,\tau}$ denote the total number of passengers exiting in the entire public transit network during time interval τ. Similarly, let $O_{C_i,\tau}$ denote the passenger outflow volume for a station cluster C_i and $O_{S_i,\tau}$ denote the passenger outflow volume for a station S_i.

Theory

Problem Definition: *Traffic In and Out Flow Prediction Problem.* At a given time interval τ and a network flow database $DB_\tau = \{tr_1, \ldots, tr_K\}$, predict the outflow of each station O_{S_i,τ_f}, cluster O_{C_i,τ_f}, and the entire network O_{E,τ_f} for a set of future time intervals τ_f.

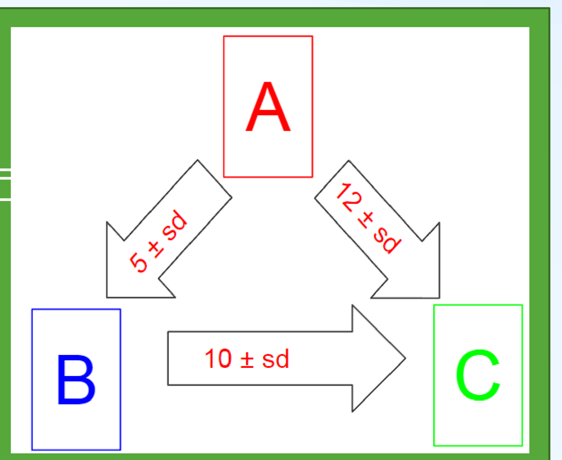
Bayes' Theorem: $P(X|Y) = \frac{P(X|Y)*P(X)}{P(Y)}$

Station Probability:

For stations A, B, and C: P(C|start = A, time ≥ 6 min.) = $\frac{P(t \geq 6, s = A|end =}{P(t \geq 6, s = A)}$

From the Theorem of Total Probability: $P(X) = \sum_{i=1}^{n} P(X|S_i) * P(S_i)$

$= \frac{P(t \geq 6, s = A|e = C)*P(C)}{P(t \geq 6|e = A)P(e=A) + P(t \geq 6|e = B)P(e=B) + P(t \geq 6|e = C)P(e=C)}$



Time Distribution Probability:

For passenger G currently in the metro network, the probability of G exiting at a given station X inn time interval τ_1

= P(G exits during τ_1| G's destination from X)

$= \frac{P(exits\ during\ \tau_1)*P(G\ did\ not\ exit \leq \tau_2|G\ exits\ during\ \tau_1)}{P(G\ did\ not\ exit \leq \tau_2)}$

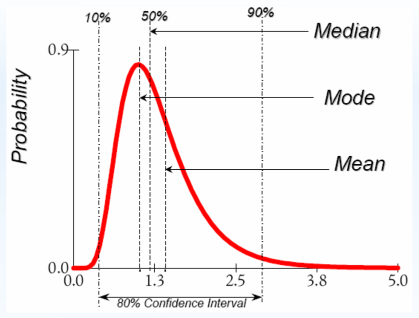
Station Bayes ✖ Time Distribution = Outflow Estimation

Methods

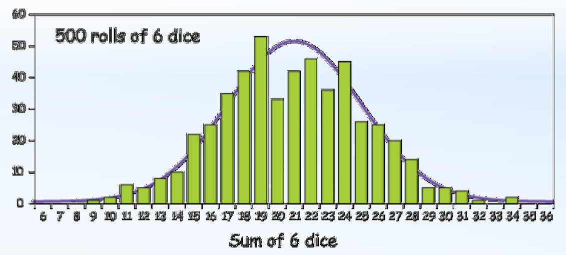
- We approach the macro-problem of traffic flow prediction from a micro standpoint by looking at each passenger
- At a given time, take all passengers traveling in the Metro network
- Predict for each passenger the probability of him at a specific time arriving at a destination station
- Partition each passenger into the micro probabilities and allocate it to each station time bucket
- Summation of the micro probabilities in a time interval for a station gives the estimated outflow
- Repeating this process for thousands of users provides an accurate prediction

Machine Learning: Gradient Boost

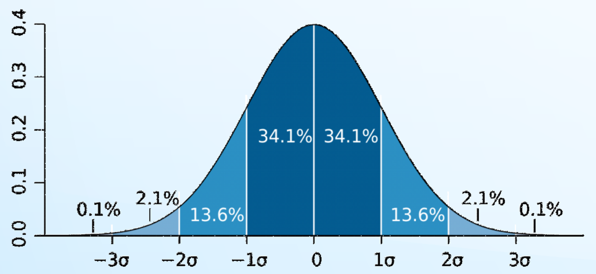
Bayes Analysis with Median



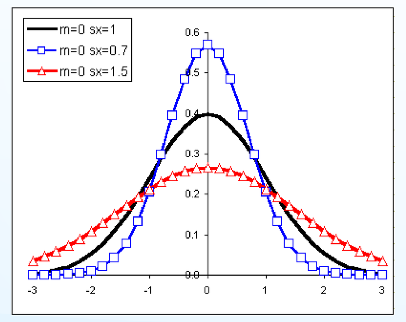
Bayes Analysis with Discrete Probabilities



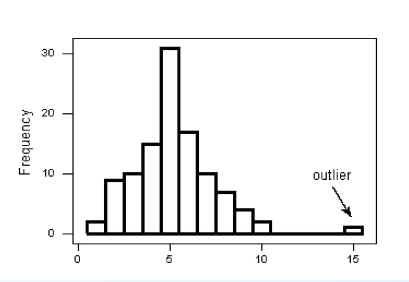
Bayes Analysis with Historical SD



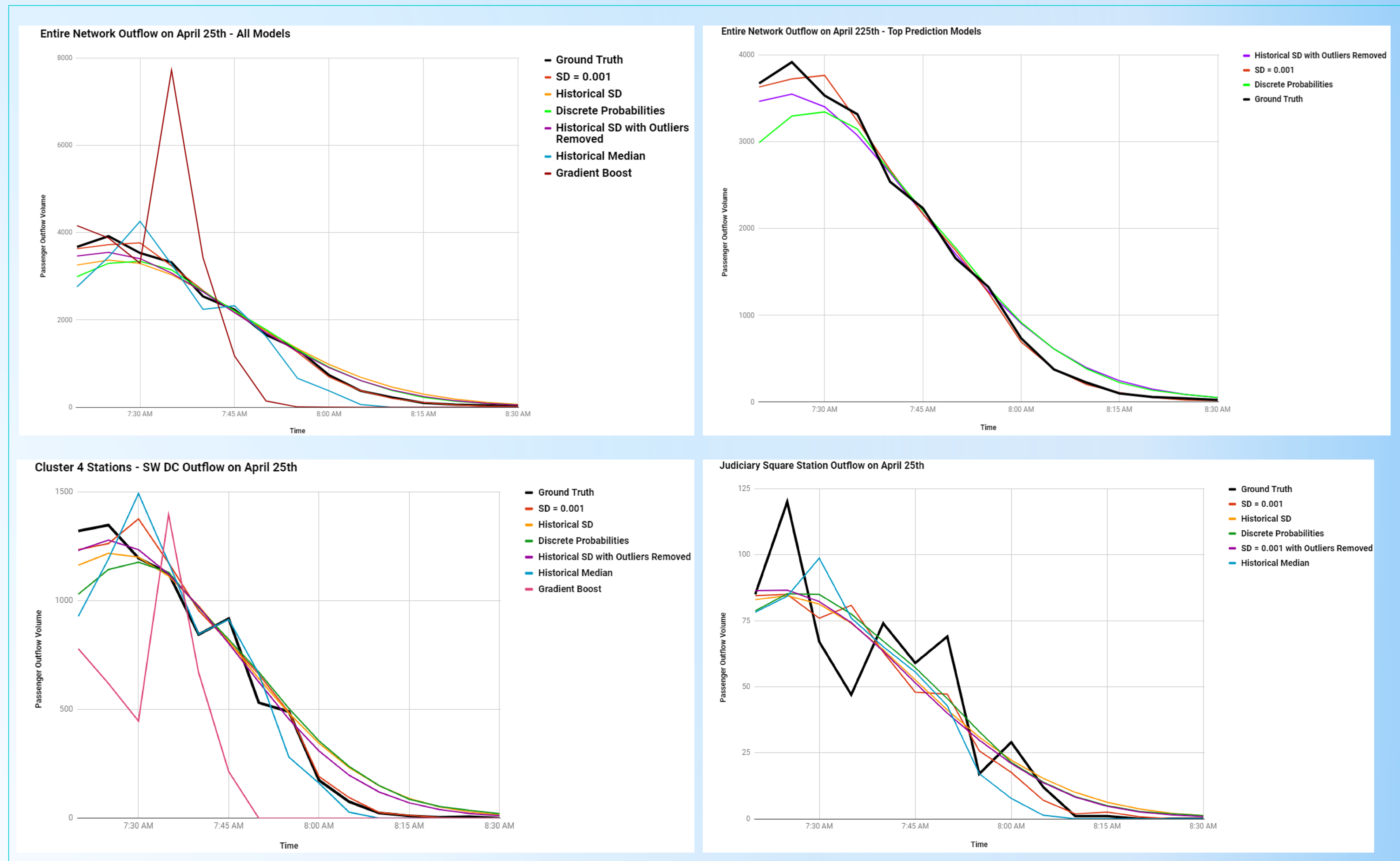
Bayes Analysis with Low SD



Bayes Analysis with Historical SD and Outliers Removed



Results



	Historical SD with Outliers Removed	SD = 0.001	Historical SD	Discrete Probabilities	Historical Median	Gradient Boost
Mean Squared Error	148.9358704	152.8131567	160.2091627	178.5028611	199.3638328	25735.62318
Root MSE	12.20392848	12.36176188	12.65737582	13.36049629	14.1196258	160.4232626

Conclusions

- The nature of metro fare card data is hard to predict
- Metro time distribution may not be normal
- Machine learning algorithms such as Gradient Boost may be overfitting training data
- Our Bayesian micro-analysis model for metro traffic flow consistently outperforms baseline approaches
- Use of Bayesian statistics can accurately predict macro traffic flow from a micro standpoint
- The Bayesian model can be easily applied to any network traffic flow with origin and destination records
- Future work may use the improved metro network flow predictions to improve road network predictions

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant CCF-1637541 titled "AitF: Collaborative Research: Modeling movement on transportation networks using uncertain data".

References

1. R. Truong, O. Gkountouna, D. Pfoser, and A. Züfle. Modeling and prediction of passenger volume in public transport. 2017.
2. Public Transportation. (n.d.). Retrieved August 04, 2017, from http://www.apta.com/mediacenter/ptbenefits/Pages/default.aspx

11th Aspiring Scientists Summer Internship Program
August 11 2017, George Mason University, Virginia