Highway Tollgates Traffic Flow Prediction

Congyang Wang, Fangyu Lin, Meng Li
Department of Computer Science, Department of Data Science – Worcester Polytechnic Institute, MA, 01609
Email: {cwang8, flin, mli6}@wpi.edu

***Abstract*—As urban population and economy grow, automobile penetration in urban area, especially metropolises, has increased fast. As a result, traffic congestion problem has arisen not only within cities, but also in suburb such as areas bridging highways. The highway is therefore no longer highway due to congestion. In this paper, our goal is to predict the average travel time from one intersection to a tollgate, as well as predict traffic volume passing through tollgates in 20-minute windows, especially during the rush hours. We are going to use ARIMA models to do the prediction of volume and travel time with data collected in the previous time window.**

***Index Terms*—Traffic Prediction; Traffic Flow Optimization; Travel Time & Traffic Volume Prediction**

I. Introduction

Urbanization has been a global trend in recent decades, since it does not only benefit rural residents by improving accessibility of urban services and resources, the local economy also has gained momentum from the process. However, various problems have been engendered as concomitants. Inevitably, living costs surges as urban population grows and overcrowding worsens living conditions. As a consequence, a growing number of citizens choose to move to satellite towns and commute on highways. The shift in lifestyle choice intensifies pressures on highway management, since increase of traffic volume may cause congestions and lower highway performance. This, as a result, entails studies on highway traffic flows.

Highway tollgates are gateways implemented for toll collection so that highway construction costs and maintenance expenditures can be reimbursed, whereas they frequently act as blockades that cause traffic congestions, especially during rush hours and public holidays, and hence incur criticism towards the administration agency. To improve highway customer satisfaction and enhance highway performance, it is necessary to solve or at least mitigate the congestion problem caused by tollgates.

Traffic volume prediction at tollgates enables administrators to make preparation in advance for digesting incoming high traffic volume, such as opening additional toll booths. Similarly, travel time prediction between tollgates and intersections empowers travelers to make choices over highway entrances, instead of running into dead-stop bumper-to-bumper situations unawarely. Moreover, with predicted traffic volume and travel time to a given tollgate, highway administration agency is able to regulate its upstream traffic flow to reduce chances of traffic congestion.

II. Related Work

There is one similar research paper [3] which evaluates the electric auto-pay system and the manual pay system. As we all know that Massachusetts highway has implemented the electric auto-pay system since 1998, which is known as "E-ZPass". However, the tollgates were not removed until 2016. Then all tollgates have been removed in Massachusetts which improved the highway traffic average speed amazingly. In this paper, we assume that there is no manual tollgates at Tollgate one, two, and three for both the entry and exit direction.

There are many good methodologies cited by other research papers. One [7] uses a loglinear model to predict the travel time in the highway. Another one [6] uses the Seasonal ARIMA Model to predict the highway volume producing more accurate results. One [4] considers the highway traffic as a multi-lane tollgates model. As we will consider the different effectiveness from the number of lanes in the tollgates, we have the data table three(links). The paper [4] considers the multi-lane as single lane to improve the average traffic velocity. Additionally, the paper [8] uses the TakagiSugenoKang Fuzzy Neural Network(TSKFNN) approach to predict the average travel time on highway and uses the back propagation neural network(BPNN) and the time series model (ARIMA) with the training trajectory data and validation trajectory data.

III. Methodology

A. Problem Setting

To study the efficiency of the highway traffic, we select a intersections and tollgates pairs as the target area. The Fig. 1 shows the detail of intersections and tollgates pairs graph. There are total six routes:

a. Routes from Intersection A to Tollgates 2 & 3;
b. Routes from Intersection B to Tollgates 1 & 3;
c. Routes from Intersection C to Tollgates 1 & 3.

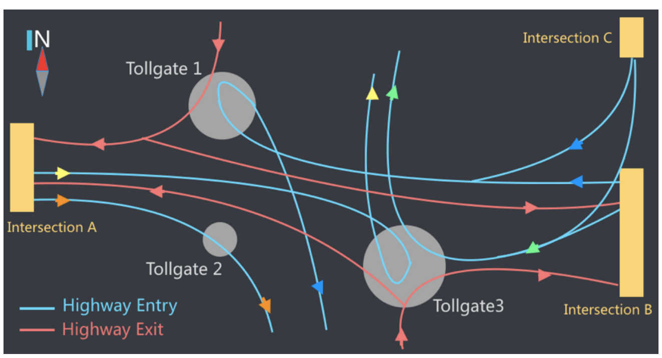

Fig. 1: Intersection and Tollgate Network Graph

The road network in Fig. 1 here used the direction graph formed by road links with interconnected. Every route in this road network is represented by a sequence of links like the model you can see in Fig. 2.

So, We divide this problem into two tasks:

Task 1: To estimate the average travel time from designated intersections to tollgates.

Task 2: To predict average tollgate traffic volume.

In the task 1, we choose the 20-minutes time window, try to estimate the average travel time of vehicles for every specific route shown in Fig. 1

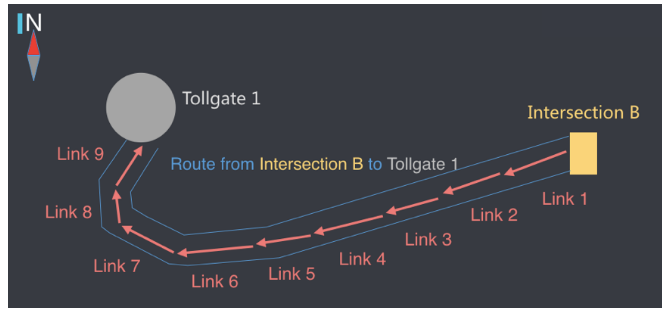

Fig. 2: Intersection and Tollgate Links Sequence

with our prediction models, so we will predict six average travel time for all routes. In the task 2, we focus on the tollgate volume prediction, we choose the 20-minutes time window as usual, try to predict the entry and exit tollgate volumes at tollgates 1,2 and 3 (tollgate 2 only has Highway Entry as shown in Fig. 1). Therefore, we will predict the volumes for the 5 tollgate-direction pairs in total.

Field	*Type*	*Description*
intersection_id	string	intersection ID
tollgate_id	string	tollgate ID
link_seq	string	a sequence of link IDs from the intersection to the tollgate separated by commas (shown in Figure 4)

Fig. 3: Routes from Intersections to Tollgates

B. Data Description

There are five datasets we choose in training process. Two for time-invariable conditions and three for time-variable factors.

The invariable data table (Fig. 3) includes the link sequence from intersections to tollgates, the graph like Fig. 2, the vehicles traveling from road intersections to highway tollgates have limited options, for each intersection-tollgate pair, this dataset selected the most important one. The data table (Fig. 4) which shows more details about infrastructure design of every piece of road link, the Fig. 5 also show the instruction of the links connection.

For the time-variable factors like vehicle trajectories data (Fig. 6), traffic volume data (Fig. 7), weather data (Fig. 8). The trajectories data includes the time-stamped records of actual vehicles along the routes in Fig. 3. The traffic volume data with the records from the three tollgates. The weather data has the features which may effect the highway traffic with 3 hour time interval.

Field	*Type*	*Description*
link_id	string	link id
length	float	length (meter)
width	float	length (meter)
lanes	int	number of lanes
in_top	string	incoming road link(s), separated by comma (as shown in Figure 3)
out_top	string	outgoing road link(s), separated by comma (as shown in Figure 3)
lane_width	float	lane width (meter)

Fig. 4: Road Link Properties

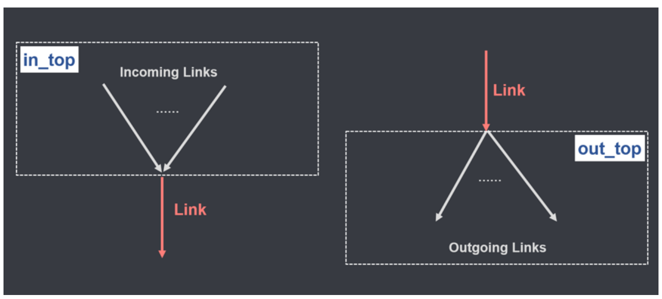

Fig. 5: Linkage In-top and Out-top Description

Field	*Type*	*Description*
intersection_id	string	intersection ID
tollgate_id	string	tollgate ID
vehicle_id	string	vehicle ID
starting_time	datetime	time point when the vehicle enters the route
travel_seq	string	trajectory in the form of a sequence of link traces separated by ";", each trace consists of link id, enter time, and travel time in seconds, separated by "#"
travel_time	float	the total time (in seconds) that the vehicle takes to travel from the intersection to the tollgate

Fig. 6: Vehicle Trajectories

Field	*Type*	*Description*
time	datetime	the time when a vehicle passes the tollgate
tollgate_id	string	ID of the tollgate
direction	string	0: entry, 1: exit
vehicle_model	int	this number ranges from 0 to 7, which indicates the capacity of the vehicle (bigger the higher)
has_etc	string	does the vehicle use ETC (Electronic Toll Collection) device? 0: No, 1: Yes
vehicle_type	string	vehicle type: 0-passenger vehicle, 1-cargo vehicle

Fig. 7: Traffic Volume through the Tollgates

Field	*Type*	*Description*
date	date	date
hour	int	hour
pressure	float	air pressure (hPa: Hundred Pa)
sea_pressure	float	sea level pressure (hPa: Hundred Pa)
wind_direction	float	wind direction (°)
wind_speed	float	wind speed (m/s)
temperature	float	temperature (℃)
rel_humidity	float	relative humidity
precipitation	float	precipitation (mm)

Fig. 8: Weather in Target Area

C. Preliminary Statistics

To explore traffic volume pattern at each tollgate, the data are partitioned into uniform time slots of 20 minutes and visualized.

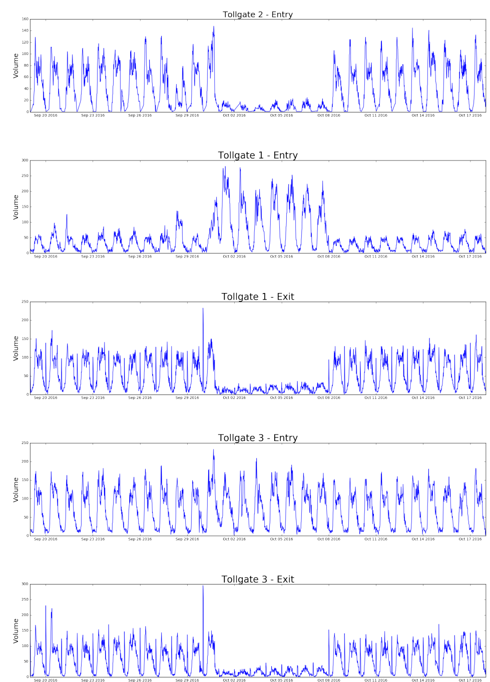

Fig. 9: Average Traffic Volume of Tollgate-Direction Pair Over Time

The Fig. 9 shows average traffic volumes in 20-minute windows versus time for each tollgate-direction pair. Apparently, average traffic volumes during October 1 to October 7 have very different patterns, compared to volumes in other time periods, except for the tollgate 3. In addition, two peaks per day are observed, in line with empirical observations that there are in general two rush-hour periods everyday.

Fig. 10 shows average travel time in 20-minute windows from intersections to tollgates. The travel time is measured in seconds. For all intersection-tollgate pairs, the average travel time reveals a pattern that is approximately random.

To figure out reasons why a large number of anomalies appear, we took a close look at some records with high travel time. We discovered that travel time of the last link of a route, i.e., the link to tollgates, dominated the overall travel time in some cases.

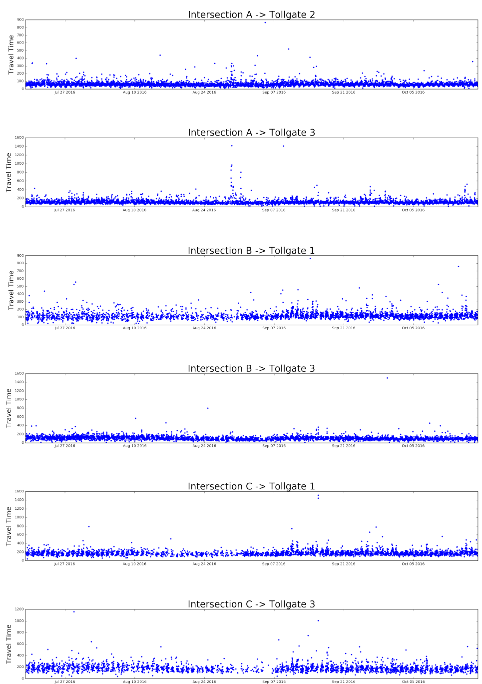

Fig. 10: Average Travel Time from Intersections to Tollgates Over Time

Then we came up with a hypothesis that the overall travel time was impacted by traffic volume at tollgates. Specifically, when traffic volume exceeds tollgate capacity, traffic congestion will arise and consequently, overall travel time from intersections to tollgates will increase because of queuing. However, in the case where traffic volume does not reach tollgate capacity, the overall travel time is not supposed to increase as traffic volume grows.

To test the hypothesis, we plotted traffic volume versus travel time in the same time window at the same tollgate. The data records selected for the plot are those with the 10% highest travel time from the intersection to the tollgate shown in plot titles, since tollgate capacities are unknown and it is unnecessary to put efforts into estimating them in the exploratory step. Any positive trend either linear or nonlinear will justify the hypothesis. However, Fig. 11 does not render any recognizable pattern, suggesting very weak relationship between traffic volume and travel time. To our surprise, high travel time appears even though traffic volume is extremely low. Therefore, the hypothesis is rejected and there must be other reasons that contribute to the anomalies and uncommon data pattern, which calls for further investigation.

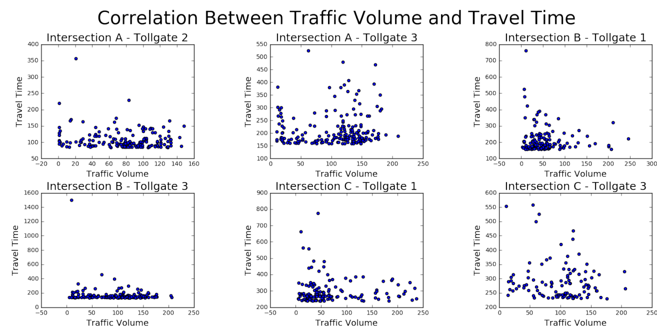

Fig. 11: Correlation Between Traffic Volume And Travel Time

IV. Experiments

A. Time Window

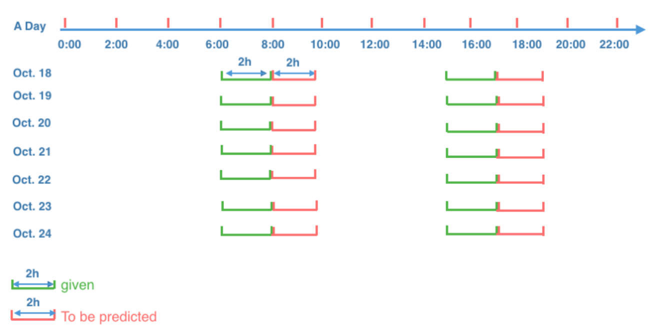

Fig. 12: Time Window

B. Evaluation Metrics

We choose Mean Absolute Percentage Error (MAPE) to evaluate the result. For task 1: We let d_{rt}

and p_{rt} be the actual and predicted average travel time for route r during time window t.

For task 2: We let C be the number of tollgate-direction pairs (as aforementioned: 1-entry, 1-exit, 2-entry, 3-entry and 3-exit), T be the number of time windows in the testing period, and f_{ct} and p_{ct} be the actual and predicted traffic volume for a specific tollgate-direction pair c during time window t.

R and T are the number of routes and number of to-predict time windows in the testing period respectively.

The MAPE for travel time & traffic volume prediction are defined as:

$$MAPE = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{T}\sum_{t=1}^{T}\left|\frac{d_{rt}-p_{rt}}{d_{rt}}\right|\right)$$

$$MAPE = \frac{1}{C}\sum_{c=1}^{C}\left(\frac{1}{T}\sum_{t=1}^{T}\left|\frac{f_{ct}-p_{ct}}{f_{ct}}\right|\right)$$

Fig. 13: Evaluation Metrics

C. Result Analysis

V. Conclusions

VI. Schedule

A. Milestone

The table below shows the timeline for this project. We are going to follow our schedule to complete each milestone on time.

Date	Milestone	Description of Work
Mar 15	Proposal Complete	2-3 Pages
Mar 22	Methodology For task-1	Methodology Due
Mar 29	Training Result Due	Training Model
April 05	Evaluation Result Due	MAPE Result
April 12	Training Result task-2	Training Model
April 19	Evaluation Result Due	MAPE Result
April 26	Final Paper Presentation	Turn In

B. Schedule

KDD CUP Timeline v.Mar.9

Week	Date	KDD	DS595	Project
Week 8	Mar/8	3/5 data released		Define tasks and get familiar with datasets
Week 9	Mar/15		3/9 Talk to Prof. Proposal Due	Phase I: work on Task 1 (model review, model selection, method develop)
Week 10	Mar/22	3/21 First Submission	Paper1 Present and Red	Phase I: data submission MAPE, model validation
Week 11	Mar/29	Daily evaluation		Phase II: work on Task 2(methods)
Week 12	Apr/5		Project Progress Presentation 12min	Phase II: Data submission MAPE
Week 13	Apr/12		Paper2 Present and Red	Phase III: Parameter Tuning
Week 14	Apr/19			Result Due
Week 15	Apr/26		P2 Presentation 22min P2 Fianl Report	Final Report Due
Week 16	May/3			
	May/10			
	May/17			
	May/24	5/25 Deadline to form teams; Data swap happens		
	May/31	6/1 Final result submission		
	Jun/7	6/5 Announcement of winners		
	Jun/14	6/15 Announcement of winners		

Fig. 14: Schedule

References

[1] G. Wu, Y. Ding, Y. Li, J. Bao, Y. Zheng, and J. Luo, Mining Spatio-Temporal Reachable Regions over Massive Trajectory Data, pp. 112, Oct. 2016.

[2] Y. Ding, Y. Li, K. Deng, H. Tan, M. Yuan, and L. M. Ni, Detecting and Analyzing Urban Regions with High Impact of Weather Change on Transport, IEEE Transactions on Big Data, vol. PP, no. 99, pp. 11, 2016.

[3] Z. Liu-Hua, C. Shi-Dong, K. Ling-Jiang, and L. Mu-Ren, The influence of tollbooths on highway traffic, 2007.

[4] K. Komada and T. Nagatani, Traffic flow through multi-lane tollbooths on a toll highway, Physica A: Statistical Mechanics and its Applications, vol. 389, no. 11, pp. 22682279, Jun. 2010.

[5] Shi Fang, K. Bian, and K. Xie, Vulnerability analysis of highway traffic networks using origin-destination tollgate data, presented at the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 2016, pp. 19571963.

[6] M. Tong and H. Xue, Highway Traffic Volume Forecasting Based on Seasonal ARIMA Model, Journal of Highway and Transportation Research and Development (English Edition), vol. 3, no. 2, pp. 109112, Dec. 2008.

[7] L. Huang and M. Barth, A Novel Loglinear Model for Freeway Travel Time Prediction, presented at the 2008 11th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2008, pp. 210215.

[8] Y. Zhang and H. Ge, Freeway Travel Time Prediction Using Takagi-Sugeno-Kang Fuzzy Neural Network, Computer-Aided Civil and Infrastructure Engineering, vol. 28, no. 8, pp. 594603, May 2013.

[9] W. Qiao, A. Haghani, C.-F. Shao, and J. Liu, Freeway path travel time prediction based on heterogeneous traffic data through nonparametric model, Journal of Intelligent Transportation Systems, vol. 20, no. 5, pp. 438448, Sep. 2015.

[10] C.-S. Li and M.-C. Chen, A data mining based approach for travel time prediction in freeway with non-recurrent congestion, Neurocomputing, vol. 133, pp. 7483, Jun. 2014.

[11] X. Xing, D. Yu, X. Tian, and Z. Cheng, Freeway travel time prediction based on clustering method with data mining, pp. 16, Sep. 2016.

[12] M. Tong and H. Xue, Highway Traffic Volume Forecasting Based on Seasonal ARIMA Model, Journal of Highway and Transportation Research and Development (English Edition), vol. 3, no. 2, pp. 109112, Dec. 2008.