

Statistical Traffic Generation Methods for Urban Traffic Simulation

Hyewon Song*, Okgee Min*

* ETRI(Electronics and Telecommunications Research Institute), 218 Gajeong-ro, Yuseong-gu, Daejeon, 34129, Korea

hwonsong@etri.re.kr, ogmin@etri.re.kr

Abstract—The urban traffic analysis is an important issue in government strategies, and there are diverse researches to address urban traffic states including congestion states. In this paper, we focus on the traffic simulation technology of various methods to analyse urban traffic states. Especially, traffic demand estimation and generation is one of key functions for simulation results to reflect real urban traffic states well. Thus, we propose the traffic demand estimation process for urban traffic simulation using trip estimation model based on L1 regularized regression model and learning the trip estimation model with trajectory data in this paper. Also, we apply the traffic demand estimation process to a case of Gangdong-gu, Seoul. Finally, we show the estimation results and simulation results by the SALT Traffic Simulator based on SUMO (Simulation of Urban MObility), so that the estimated trips are similar to real traffic patterns and the simulation results from estimated trips is within about 10% errors coverage.

Keywords— Urban Traffic Simulation, Urban Traffic Generation Model, Traffic Demand (Trip) Estimation, Regularized Regression Models

I. INTRODUCTION

Recently, it is considered as an import society welfare issue to provide convenient and comfortable transportation systems for urban people, and there are many researches to address urban traffic states and to solve problems caused by the issues from urban traffic states using smart and intelligent transportation systems. One of them is urban traffic simulation technology. [2]-[8] The traffic simulation technology is more powerful in a case of problems not to be applicable to real worlds immediately, for example, the analysis of effectiveness according to changing traffic signal strategies can be performed using the traffic simulation. For this traffic simulation technologies, the traffic demand estimation, especially traffic demand estimation and generation methods which reflects real traffic states well, is one of basic and main functions. Thus, there are various researches in this area, and recently some of them focus the machine learning based statistical methods or deep learning based methods. [9]-[11]

In this paper, we propose the traffic demand estimation and generation methods using L2 Logistic regression model of machine learning methods. In this process, we use trajectory data collected by navigation systems and traffic census data from Seoul Topis [12] as training and testing the model. Using this learned model, we estimate and generate traffic demands

data for a traffic simulation generated as a xml format, which is used in traffic simulator developed by SALT (Simulation for Analyzing Loads in Traffic) project using an open source traffic simulator, SUMO (Simulation of Urban MObility) [13]. Moreover, this estimated traffic demand is provided as one of input data of the traffic simulator. Finally, we show the estimated traffic pattern is similar to real traffic pattern throughout the results from the simulation.

II. STATISTICAL TRAFFIC GENERATION PROCESS FOR URBAN TRAFFIC SIMULATION

In order to infer the status of urban traffic as an input data for the urban traffic simulation, we focus on a statistical traffic generation model with trajectory data. As one of statistical methods, we consider the logistic regression model with a L2 regularization. Based on regression model, the traffic generation model is learned and select the origin and destination for trips as an input data of the traffic simulation.

Based on this idea, we propose the Statistical Traffic Demand Generation (STDG) process with the L1 regularized logistic regression model in this paper. This statistical traffic demand generation process contains 3 steps: 1) Data cleaning and preprocessing, 2) Trip estimation based on machine learning model, 3) Trip assignment using SUMO. Figure 1 shows the conceptual construction and functional flow of the proposed STDG process.

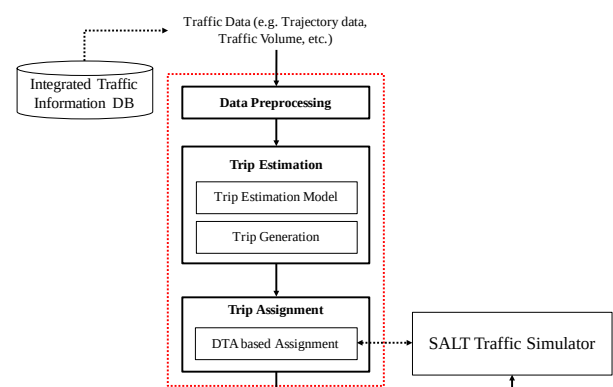


Figure 1. Conceptual Construction and Functional Flow

As shown in Figure 1, the collected traffic data such as trajectory data, traffic census data, and so on, is saved in the Integrated Traffic Information DB designed by SALT project

and it can be provided for data cleaning and preprocessing in the Data Preprocessing step in the STDG process. And then, using preprocessed data, the L2 regularized logistic model based trip estimation model is learned, so that, the estimated trip can be generated in the Trip Estimation step in the STDG process. The generated trip is assigned using DTA (Dynamic Traffic Assignment) [14] based trip assignment method, and the result is finally provided to the SALT Traffic Simulator based on SUMO as one of input data.

III. TRAFFIC ESTIMATION AND GENERATION WITH GANGDONG-GU CASE

A. Data Cleansing and Preprocessing

In this paper, we mainly deal with trajectory data and traffic census data and they are used directly or indirectly in the trip estimation process. In a case of trajectory data, it is based on map data such as shape file, so that, we suppose the traffic demands can be represented using a trip with a pair of origin edge and destination edge in this paper. Also, it is provided by a matrix shape like a O/D matrix. Under this supposition, we perform data cleaning and preprocess. In this step, the data is filtered by target area and target time, and drop the unsatisfactory data for extracting trip information, for example, data without a pair of a start edge and an end edge.

Especially, the trajectory data is used as a key information for inference of trip patterns in this paper. As mentioned in [11], the trajectory data is a bias data for more accurate and predictable trip generation based on learning process using machine learning model like a regression model.

As a trajectory data, we use the 7-day-data (from Jun. 1, 2017 to Jun. 7, 2017) within limited areas is covered with 4 administrative districts, Gangdong-gu, Gangnam-gu, Songpa-gu and Seocho-gu. In this paper, we mainly target to Gangdong-gu, and extract trip information including start and end pair and intermediate links. In this pre-processing, we refer to OD matrix data from [15] and traffic volume data from [12] related to the target area. The data patterns of preprocessing trajectory data, traffic volume data, OD matrix data are shown Figure 1, Figure 2, and Figure 3, respectively.

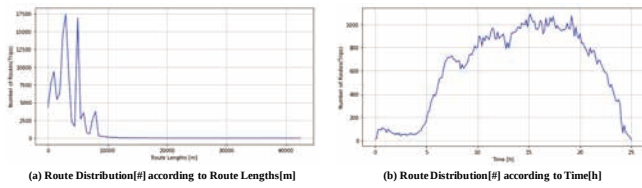


Figure 2. Preprocessing Trajectory Data (Route/Trip Distribution Pattern)

Figure 2 means the route distribution of the trajectory data as the number of routes according to route lengths (Figure 2(a)) and time (Figure 2(b)), respectively. Figure 3 represent the collected traffic volumes in traffic census spots and the label name in the graph (from the B-09 to the C-22) is refer to Topis [12]. In this paper, we mainly refer the data from B-09 and B-10 because they located in boundary loads of Gangdong-gu.

Figure 4 show a heat map graph based on the O/D matrix from [17]. In Figure 4, the x-label and y-label means an identification number of TAZ (Traffic Analysis Zone) belong in Gangdong-gu, Seoul and the colors are deeper and deeper according to the number of trip occurrence between a TAZ on x-axis (Origin TAZ) and a TAZ on y-axis (Destination TAZ).

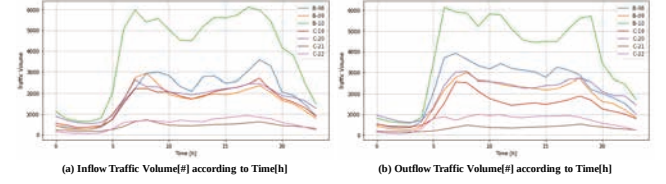


Figure 3. Preprocessing Traffic Volume per 1 hour

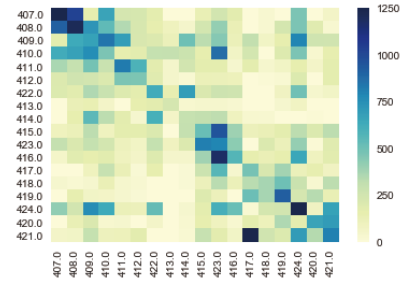


Figure 4. Preprocessing OD matrix per 1 day

B. Trip Estimation

In the STDG process, the Trip Estimation step is divided by 2 parts: Estimation and Generation. In the Estimation part, we suppose that the trip with a pair of a start edge and an end edge can be selected with probability inferred from historical trip information. Under this supposition, we derive the model to infer the value matrices of selected start edges and end edges with time varying property individually.

$$\begin{aligned} x_{s_i}(t_k) &= P[n_{s_i}(t_k)|w_{s_i}(t_{k-1})] \\ x_{e_i}(t_k) &= P[n_{e_i}(t_k)|w_{e_i}(t_{k-1})] \end{aligned} \quad (1)$$

Eq. (1) defines the member values for the value matrix of selected start edges (x_s) and the value matrix of selected end edges (x_e). It has time varying property, so that, we consider the discrete time interval. Namely, t_k is the k -th time slot and t_{k-1} is the $(k-1)$ -th time slot s_i means the i -th edge of start edges and e_i means the i -th edge of end edges. Also, n_s and n_e are the number of occurrence for a specific edge to be selected as a start edge and an end edge, individually. Moreover, w_s and w_e are the value in previous time slot (t_{k-1}) for a specific edge to be selected as a start edge and an end edge, individually.

When the number of total edges belong in Gangdong-gu, Seoul is N , we respectively calculate the $(N \times 1)$ value matrices of start edges and end edges based on Eq. (1) using preprocessed trajectory data. And then, it can be used as a training data set (6 days) and a testing data set (1 days) for learning the estimation model based the L1 regularized logistic regression model. Finally, we generate the trip data for 1 day (24h) using the trip estimation model of the Estimation part in

the Generation part. The Table 1 represents the comparison of relative performance values, which means the mean squared error (MSE), the mean squared log error (MSLE) and the R square (R2), between ground truth value given by real data and the generated trip data.

TABLE 1. RELATIVE PERFORMANCE VALUES

	MSE	MSLE	R2
Start (Origin) Edges	2.24	0.13	0.47
End (Destination) Edges	2.41	0.17	0.34

As shown Table 1, we calculate the MSE, MSLE and R2 using the weight value vectors for start edges and end edges respectively. This weight value vectors represent the number and probability of selected edges as a start edge or an end edge. Also, for this weight value vectors, we ignore edges which were not selected in the Generation part because the number of edges being not selected are larger than the number of edges being selected.

C. Trip Assignment

In this paper, we use the SALT Traffic Simulator based on SUMO, which is an open source traffic simulator, and the SUMO provides various methods for traffic demand generation [13]. The DUA (Dynamic User Assignment) Routing is one of them, and it is enable to provide route data with the user equilibrium based on DTA (Dynamic Traffic Assignment) Scheme. Thus, in this paper, we use the DUA Routing for the Trip Assignment step.

Using the estimated trip from the previous step, we generate the route data as an input data for SALT Traffic Simulator. Figure 5 shows the route distribution of generated route data as the number of routes according to route lengths and time, individually. In Figure 5, a blue line means original route distribution of trajectory data and a red line means route distribution of generated route data in the Trip Assignment step. As shown in Figure 5, the number of the generated routes are about 2.8~3.0 times the number of routes given by original trajectory data. Also, it is similar to the pattern and amount of traffic volumes in Figure 3.

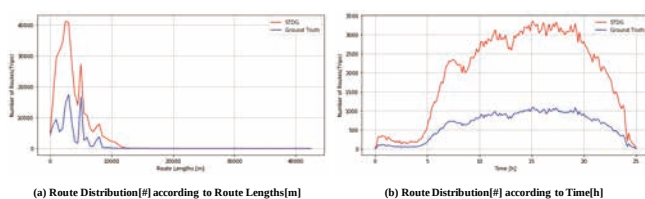


Figure 5. Route Distribution[#] of Route Data generated after the Trip Assignment step

IV. EXPERIMENTS AND RESULTS WITH GANGDONG-GU CASE

According to the STDG process, we generate route data, which is the traffic demand data for the SALT Traffic Simulator and is estimated as 1-weekday-trip-data for Gangdong-gu, Seoul. Using it, we simulate the traffic status of Gangdong-gu for 1 hour and 24 hours when it is a common weekday.



Figure 6. Road Network of Gangdong-gu, Seoul

Figure 6 shows the road network of Gangdong-gu for the traffic simulation. This road network consists of edges based on the specific navigation map and TLS-programs based on Traffic Light System (TLS) of Seoul [12].

As one result from the simulation, we calculate the travel time according to a time interval (5min). Figure 7 shows the travel time pattern for 1 hour per 5 min. Also, Figure 8 shows the travel time pattern for 24 hours per 30 min. Both of them, a blue line means ground truth and a red line means simulated value.

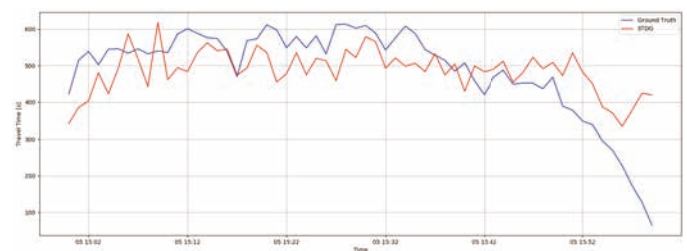


Figure 7. Travel Time (for 1 hour) according to time interval(= 5 min)

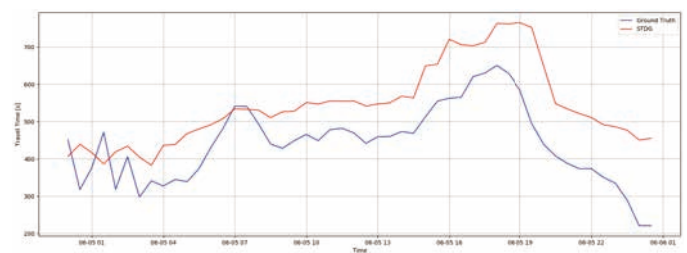


Figure 8. Travel Time (for 24 hour) according to time interval(= 30 min)

As shown in Figure 7, the differences between ground truth and simulated value is within about 8-9% coverage, although there are larger than 10% in some time slots of a fore part and a back part. However, as shown in Figure 8, the differences between ground truth and simulated value is about 22% averagely. As one reason for this result, we guess our SALT simulator is implemented for the PoC (Proof of Concept), so that the capacity is not sufficient to cover whole inferred traffic volume yet.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose the STDG process based on L2 regularized logistic regression model, and then, we estimate and generate trips as traffic demands for Gangdong-gu, Seoul, using the proposed STDG process. Also, we show the estimated trips after learning process with a training data set and simulation results with the estimated trips, and we confirm the proposed STDG process performs within a reasonable coverage. However, we just use the trajectory data for 7 days, and current SALT Traffic Simulator is developed as PoC (Proof of Concept) level. Thus, we should improve the STDG process with plenty data (at least 1 year) and SALT Traffic Simulator with well-defined functions.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2017-0-00121, Development of a Traffic Predictive Simulation SW for Improving the Urban Traffic Congestion).

REFERENCES

- [1] G. Papageorgiou and A. Maimaris, "Modelling, Simulation Methods for Intelligent Transportation Systems," *Intelligent Transportation Systems*, Dr. Ahmed Abdel-Rahim (Ed.), InTech, 2012.
- [2] S. M. Metev and V. P. Veiko, "Urban Traffic Simulation with SUMO," 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [3] J. C. Dias, P. H. Abreu, D. C. Silva, G. Fernandes, P. Machado and A. Leitao, "Preparing Data for Urban Traffic Simulation using SUMO," *The SUMO User Conference (SUMO2013)*, May, 2013
- [4] T. Suzumura, G. McArdle and H. Kanezashi, "A high performance multi-modal traffic simulation platform and its case study with the dublin city," *Proc. of the 2015 Winter Simulation Conference (WSC15)*, p. 767-778, Dec. 2015.
- [5] T. Toledo, M. E. Ben-Akiva, D. Darda, M. Jha, and H. N. Koutsopoulos, "Calibration of Microscopic Traffic Simulation Models with Aggregate Data," *Journal of the Transportation Research Board*, No. 1876, pp. 10-19, 2004
- [6] E. F. Z. Santana, N. Lago, F. Kon and D. S. Milojicic, "InterSCSimulator: Large-Scale Traffic Simulation in Smart Cities using Erlang," 18th *Workshop on Multi-agent-based Simulation (MABS2017)*, May, 2017
- [7] R. Balakrishna, D. Morgan, A. Rabinowicz and H. Slavin, "Advances in Origin-Destination Trip Table Estimation for Transportation Planning and Traffic Simulation," Association for European Transport and Contributors, 2008.
- [8] M. E. Ben-Akiva, S. Gao, Z. Wei and Y. Wen, "A dynamic traffic assignment model for highly congested urban networks," *Transportation Research Part C: Emerging Technologies*, vol. 24, p. 62-82, 2012
- [9] S. Bera and K.V.K. Rao, "Estimation of origin-destination matrix from traffic counts: the state of the art," *European Transport*, no.49, p. 3-23, 2011
- [10] T. Khan and M. Anderson, "Accurately Estimating Origin/Destination Matrices in Situations with Limited Traffic Counts: Case Study Huntsville, AL," *Journal of Traffic and Transportation Engineering*, vol.5, no. 3, p. 64-72, 2016
- [11] T. Morimura and S. Kato, "Statistical Origin-Destination Generation with Multiple Sources," *Proc. of International Conf. on Pattern Recognition (ICPR 2012)*, p 3443-3446, Nov. 2012
- [12] Topis (Seoul Transport Operation & Information Service). [Online]. Available: <http://topis.seoul.go.kr/eng/english.jsp>
- [13] SUMO Traffic Demand Modelling. [Online]. Available: http://sumo.dlr.de/wiki/Demand/Introduction_to_demand_modelling_in_SUMO
- [14] *Dynamic Traffic Assignment: A Primer*, Transportation Research Circular No. E-C153, Transportation Research Board, 2011
- [15] Korea Transport DataBase. [Online]. Available: <https://www.ktdb.go.kr/eng/index.do>



Hyewon Song received a degree of MS in 2005 and had been a PhD candidate since 2005 in Korea Advanced Institute of Science and Technology (KAIST). She is currently a senior researcher and has worked at Electronics and Telecommunications Research Institute (ETRI) since 2010. Recently, she

joins to the project for development of traffic simulator based on traffic data collected by multi-sources in Seoul, and interests in data analysis using machine learning mechanism including deep learning.



Okgee Min received her BS and MS degrees in Computer Science and Statistics from Chung-Nam National University, Taejeon, Korea in 1988 and 1992, respectively. She received her PhD in Computer Engineering from Chung-Nam National University in 2010. She is currently a director and principal

researcher at ETRI. Her research interests include data stream processing, big-data analysis, data intelligence, and cognitive architectures.