Technology Reports Real-time Population Statistics Congestion Prediction Machine Learning

Special Articles on AI Supporting a Prosperous and Diverse Society

Avoiding Tokyo Bay Aqua Line Congestion Using Traffic Congestion Forecasting AI —Prediction Based on Statistical Processing of Mobile Phone Network Operations Data—

Research Laboratories Masayuki Terada Hiroto Akatsuka
Platform Business Department Tomohiro Nagata
Corporate Sales and Marketing Department I Satoshi Nakanishi

Traffic Congestion Forecasting AI is a technology that can predict the occurrence of traffic congestion, including size and times, by applying AI technology to real-time population statistics that are created in near-real-time from mobile telephone network operations data. Predictions are made based on the number of people out on a given day, making accurate predictions possible, accounting for effects such as weather or special events. This article gives an overview of Real-time Population Statistics and Traffic Congestion Forecasting AI, and introduces a trial performed in cooperation with NEXCO East on the Tokyo Bay Aqua Line. An overview of the trial, evaluation of prediction accuracy, and results of a survey of users participating in the trial are discussed.

1. Introduction

Frequent traffic congestion has been a major issue for many years in Japan. The resulting economic losses have been estimated at over 10 trillion yen per year [1], exerting a strong negative pressure on economic activity. Beyond effects on the economy, we are also all familiar with the related decrease in the quality of daily life. As an example, a common experience is encountering a traffic jam on the way home from an outing on the weekend and how this can diminish the enjoyable memories of the event.

While the roads that tend to get congested and

* Real-time Population Statistics: A service providing mobile-network spatial statistics in near real time. Displays aggregate population by area and attributes but does not include identifying information. These population statistics are created according to the mobile Spatial Statistics Guidelines.

the days and times are generally known, there can be large differences in day-to-day conditions: whether congestion will occur, how large it will be, when it will start, and when it will clear. For example, routes home from popular tourist destinations often get congested. One can even encounter an unexpected large traffic jam in an area that does not normally get congested if large numbers of people gather on that day for an event or other reason. On the other hand, if the weather is bad and people stay home, congestion could decrease or not occur at all.

As such, knowledge of approximately how many people are actually out on that day (turnout) is needed to predict what will actually happen in the during times when people are returning home, including the areas around their destinations. If we can know the turnout on a day quantitatively, we can expect to be able to predict congestion that will occur during a return-home period.

Traffic Congestion Forecasting AI is a new traffic prediction technology developed with a focus on this relationship between turnout and congestion. It is able to comprehend turnout on a given day using Real-time Population Statistics, a technology that estimates human populations throughout Japan using operations data from the NTT DOCOMO mobile telephone network, and based on this information, it is able to predict congestion and the scale and time-frame of the congestion.

This article gives an overview of Real-time Population Statistics and Traffic Congestion Forecasting AI and describes tests conducted in collaboration with East Nippon Expressway Co. Ltd. (NEXCO East) on the Tokyo Bay Aqua Line expressway starting in December 2017.

2. Real-time Population Statistics

Real-time Population Statistics is a new form of population statistics arising from R&D to make a real-time version of Mobile Spatial Statistics*1 [2], a commercial service that has been offered since 2013. It is able to estimate population distributions throughout Japan on a 500 m grid*2 (with some areas such as centers of designated cities on a 250 m grid) according to attributes such as age group (in 5-year increments) and place of residence (city, town, etc.). Data fluctuations can be provided at 10-minute intervals, approximately 20 minutes after the fact. In other words, population distributions at 12:00 are available by 12:20, those at 12:10 are available by 12:30, and so on.

An example visualization of Real-time Population Statistics is shown in **Figure 1**. This is an illustration of a population distribution at noon on a weekday. Each grid section is colored according to the population density in the cell, using blue, green, yellow, and red, in order of increasing population density.

Figure 2 shows the number of visitors to the Sumida River Fireworks Display on a given year at 8 pm, on a 500 m grid (defining the number of visitors as the increase in population compared with the usual population). Here, the number of visitors is represented by red, with darker shades indicating more visitors in that grid section.

The red areas are concentrated along the Sumida River, showing that many spectators were gathered there, but there are also two areas slightly east of the river with concentrations of people. Investigation on the following day revealed that, although they are somewhat far from the river, these

*1 Mobile Spatial Statistics: Population statistical data generated according to the "Mobile Spatial Statistics Guidelines," from NTT DOCOMO mobile network operations data. Population distributions on a grid (see *2) and by municipal boundaries are estimated such that individual users cannot be identified, using an estimation of the number of mobile phones currently in each base-station area and adjusting based on base-station area data, NTT DOCOMO phone usage rates and other information.

*2 Grid: Land divided into sections based on latitude and longitude.

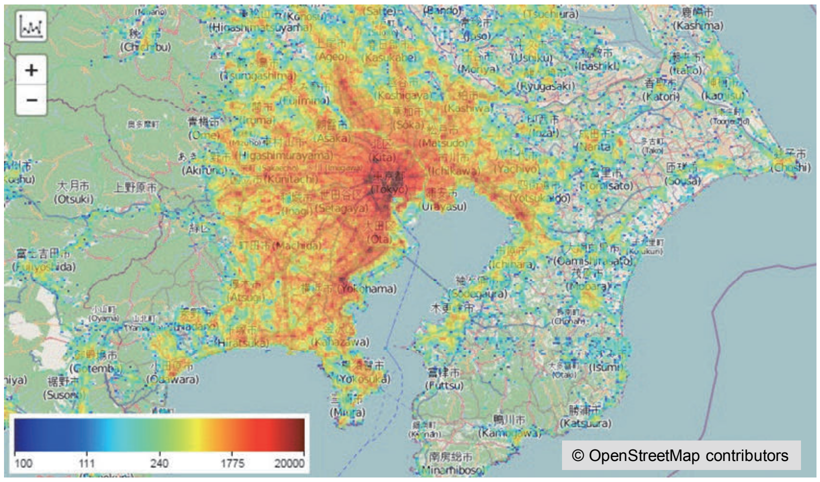

Figure 1　Real-time Population Statistics example

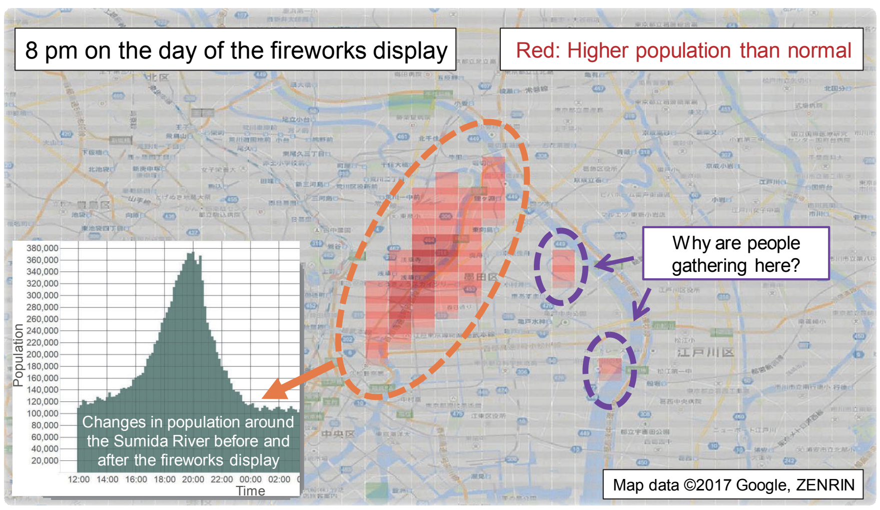

Figure 2　Results of estimating number of visitors to the Sumida River Fireworks Display in a given year

are little-known spots that are good for viewing the fireworks with few obstructions.

This illustrates the strength of Real-time Population Statistics, providing quantitative data on the fluctuations in population density throughout Japan according to attributes such as age and place of residence in near-real-time, including gatherings of people in areas, even without knowing why they may be gathering there.

3. Predicting the Future Based on Real-time Population Statistics

Real-time Population Statistics makes it possible to know the dynamics of population distributions or changes in how people are gathering. This suggests that dynamics of social phenomena and economic trends that are correlated to the movements of people can be estimated from Real-time Population Statistics.

A correlation is a relationship in which, when one value increases or decreases, another value also increases or decreases. For example, when the temperature increases in the summer, sales of ice cream also increase. Conversely, sales decrease when it is cool. We say there is a correlation between temperature and ice cream sales. By calculating this relationship based on past temperature fluctuations and ice cream sales records, one of the values can generally be estimated if the other value of is known.

In another example, water levels in rivers rise when it rains and drop after the rain stops. As with this example of precipitation and water levels, some correlations also involve a time difference. The value that changes first is called the leading indicator, and the value that changes later is called the lagging indicator. Since precipitation changes before the water levels, precipitation is a leading indicator for water levels. If the correlation can be computed based on past fluctuations in precipitation (considering time differences) and water levels in rivers, it will be possible to predict future fluctuations in water levels from precipitation leading up to the present. In other words, the leading index has the potential to predict future values of the lagging indicator.

Accordingly, changes in phenomena that are correlated to population fluctuations can be estimated based on Real-time Population Statistics, even if they cannot be observed directly. Human behavior involves a wide range of social and economic activities, and various social phenomena and economic trends are correlated to the movements of people. In particular, these correlations can also have a time difference, as with the correlation between precipitation and river water levels, so when population is the leading indicator, future changes in the social phenomena may be predictable.

4. Traffic Congestion Forecasting AI

Traffic conditions are an example where the future can be predicted. In implementing Traffic Congestion Forecasting AI, NTT DOCOMO has focused on increases and decreases in population in a given area as a leading indicator for increases and decreases in traffic demand on routes taken to return home from that area. The system is able to accurately predict traffic conditions several hours later, which has conventionally been difficult. This is done by making predictions based on observed population distributions, which are the basis of traffic demand. Thus, it can predict changes in traffic conditions from the afternoon until late at night based on populations observed at mid-day and earlier.

This gives users an opportunity to check traffic forecasts after lunch and revise plans for going home in the afternoon based on the information. Thus, it can help reduce the misfortune of encountering a traffic jam on the way home and having enjoyment ruined, as touched on earlier in this article. If an increasing number of people act to avoid the congestion based on predictions, traffic demand will diffuse over time, relaxing or even eliminating the congestion itself. If people also avoid the congested times by deciding to have dinner before going home, for example, they will also spend more time or money in the area. In these ways, changes in user behavior based on prediction information can be expected to mitigate congestion by distributing traffic and stimulate economic activity in surrounding areas.

4.1 Technical Overview

The traffic predictions from Traffic Congestion Forecasting AI are implemented using population distributions obtained from Real-time Population Statistics and by applying a type of AI technology called machine learning*3. Specifically, a congestion prediction model that formulates the relationship between population and traffic conditions is created by training it using data from a set period in the past, consisting of population distributions together with the traffic history for the same day. When making predictions each day, population distributions from noon that day are presented to the congestion prediction model to obtain results predicting traffic conditions during a return-home period. This is presented schematically in **Figure 3**.

Here, we want to stress that by population, we do not mean simply the number of people in the given area, but rather, the population distribution on the grid, by attribute, as obtained from Real-time Population Statistics.

As an example, we consider congestion predictions on Tokyo Bay Aqua Line, an expressway crossing the Tokyo Bay between Kawasaki City in Kanagawa Prefecture and Kisarazu City in Chiba Prefecture. It is also the subject of a trial described below.

If the turnout on the Boso Peninsula is large

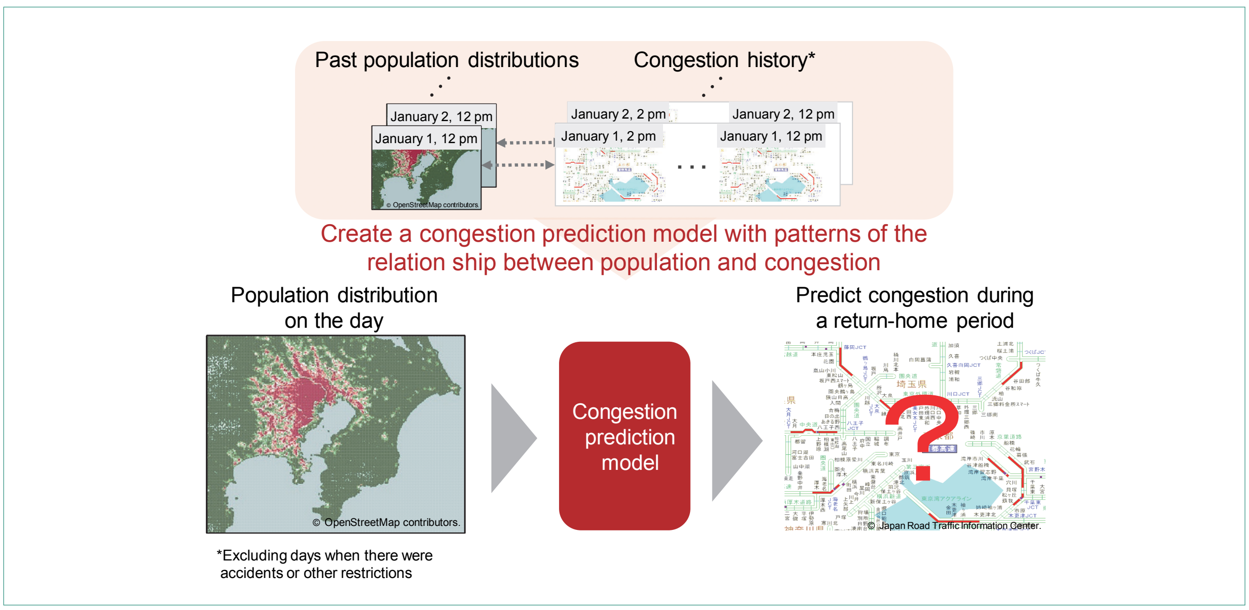

Figure 3　Traffic Congestion Forecasting AI organization

*3　Machine learning: A framework that enables a computer to learn the relationships between inputs and outputs by statistical processing of examples.

and mainly of people living in Chiba, there is almost no effect on congestion on the Aqua Line. On the other hand, if there are many people from Tokyo and Kanagawa, congestion on the Aqua Line during a return-home period can be severe. Even within Tokyo, whether people are coming from the east or the west of Tokyo can greatly affect traffic congestion.

Effects can also differ greatly depending on where visitors stay on the Boso Peninsula. Most people in the north of the peninsula will use the Keiyo Expressway or other routes, and people in the south end may use the Tokyo Bay Ferry. These proportions also differ depending on where they are going. For example, a larger proportion of people going to a golf course will be travelling by car, compared with other destinations, and they will tend to arrive earlier in the morning and go home earlier. As such, we can expect they will contribute to increasing traffic demand earlier in the day.

In this way, the effects of population distributions on congestion are not determined simply by the total number of people but can differ greatly according to location and other attributes. Traffic Congestion Forecasting AI is able to make predictions incorporating such differences by using AI to formulate such effects of population distributions by attribute on congestion, obtained from Real-time Population Statistics.

4.2 Tokyo Bay Aqua Line Trial

As part of implementation trials verifying the utility of Traffic Congestion Forecasting AI, we conducted a trial in collaboration with NEXCO East on the Tokyo Bay Aqua Line starting in December 2017 [3].

In this trial, we used population distributions at noon on the Boso Peninsula to predict congestion on the Kawasaki-bound lanes of the Tokyo Bay Aqua Line, which often occurs on weekend evenings and into the night. In particular, we predicted whether congestion would occur during the period from 14:00 to 24:00 based on population distributions by attribute at 12:00 in the Boso area, including residential areas. When congestion was predicted, we also predicted the start and end times, the peak time, and the physical length of the congestion at the peak time. In December 2018, we implemented new methods based on customer survey results for predicting the time required to travel the length of the congestion and the traffic demand every 30 minutes over the same time period. The survey results and new methods are described in detail below.

The results predicted by Traffic Congestion Forecasting AI are provided every day to the driving public through the Drive Plaza Web site operated by NEXCO East, which provides information on expressways in Japan. During the trial, in addition to the congestion prediction results, coupons were also issued, offering discount for meals and shopping at Kisarazu and other locations (called "Yorutoku coupons"). This was intended to mitigate congestion by spreading the return traffic over time and to stimulate local economies. An overview of the trial is shown in **Figure 4**.

4.3 Evaluation of Prediction Accuracy

Before providing prediction information to the public, we evaluated the accuracy of Traffic Congestion Forecasting AI. The evaluation was conducted over two years and four months, between

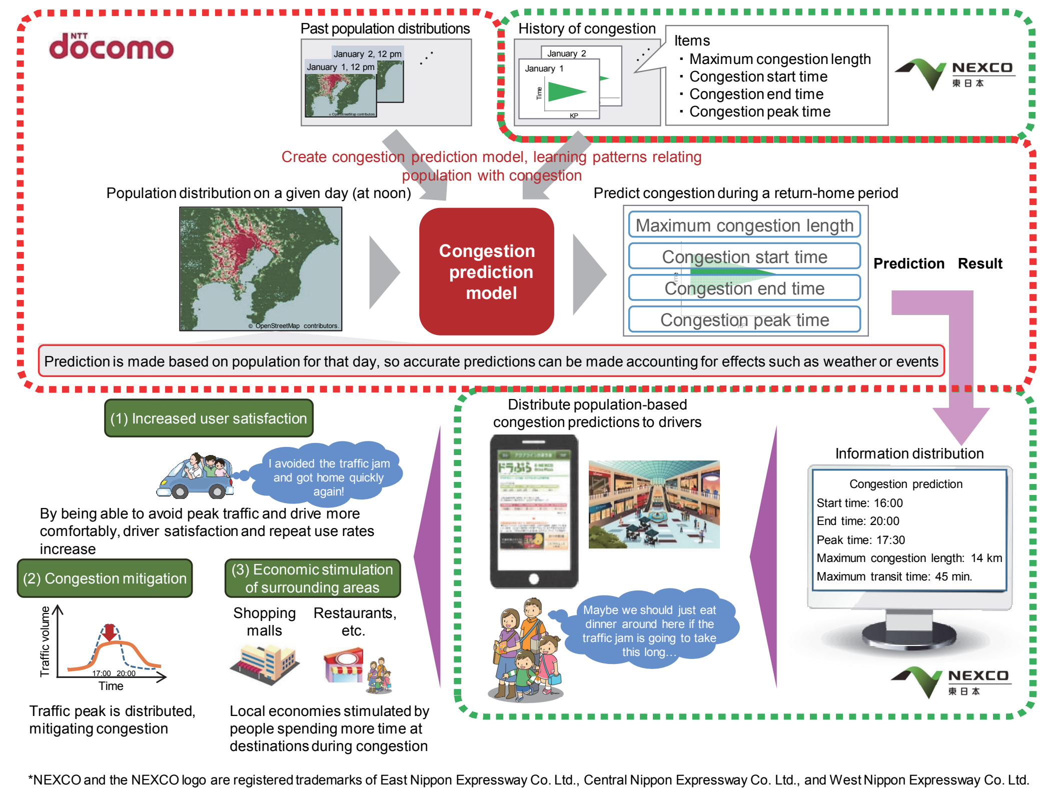

Figure 4 Overview of collaboration with NEXCO East

January 2015 and April 2017 (excluding days with accidents and other traffic restrictions), and congestion was predicted on each day during the period using Leave-One-Out Cross Validation (LOOCV)*4. Ground truth data used for training and examination consisted of traffic history data maintained by NEXCO East for the time of the trial.

We used two indices for evaluation: the Missed-Alarm Rate (MAR) is the number of days congestion occurred even though it was predicted not to occur divided by the total number of days congestion did occur. The False-Alarm Rate (FAR) is the number of days when congestion did not occur even though it was predicted to occur divided by the total number of days congestion was predicted to occur. The accuracy of Traffic Congestion Forecasting AI is shown in **Table 1**, compared with results from "Congestion Forecast Calendar," which has been provided earlier by NEXCO East, as a benchmark.

As an example, compared with results from Congestion Forecast Calendar, the MAR went from 6% to 1% and FAR from 18% to 0% for congestion longer than 10 km. These can both be considered great improvements. For the FAR in particular, the results were improved overall. Note that only population data, and no other information (day of

*4 LOOCV: A method for evaluating the accuracy of a statistical predictor.

the week, weather, event information, etc.), was used for these predictions. These results were obtained by giving only the population distributions from Real-time Population Statistics for that day to the congestion prediction model trained with past data.

4.4 Survey Results and Introduction of a New Method

As part of the joint trial, a Web survey regarding the trial was conducted during the period from March 20 to July 9, 2018 [4]. The survey was completed by people who agreed to participate after learning of the survey through pamphlets placed in tourism facilities on the Tokyo Bay Aqua Line and in Chiba Prefecture, e-mails distributed to users of the Drive Plaza*5 and Drive Traffic*6 Web sites, and banner advertisements. Excerpts of the results are shown in **Table 2** and **Figure 5**.

Over 90% of the respondents to the survey indicated an intention to use the service in the future. In particular, approximately 95% of respondents presumed likely to use the Aqua Line frequently (those living outside of Chiba prefecture in the Kanto area, using it for leisure more than once every six months) indicated an intention to use it in the future.

The survey also confirmed a strong demand for information to be provided by time period as a desired feature of Traffic Congestion Forecasting AI in the future. Given this intent to use the service and requests for features, we developed new technology for predicting at 30-minute intervals the time needed to traverse the Aqua Line and traffic demand. We then updated the pages providing Traffic Congestion Forecasting AI information on the

Table 1 Evaluating accuracy of Traffic Congestion Forecasting AI

(a) Missed-alarm rates for actual congestion length

Congestion length	Missed-alarm rate	
	Congestion Forecast Calender	Traffic Congestion Forecasting AI
15 km and greater	2%	0%
10 km and greater	6%	1%
5 km and greater	7%	3%

(b) False-alarm rates for predicted congestion length

Congestion length	False-alarm rate	
	Congestion Forecast Calender	Traffic Congestion Forecasting AI
15 km and greater	6%	0%
10 km and greater	18%	0%
5 km and greater	22%	6%

Table 2 Intention to use Traffic Congestion Forecasting AI in the future

	Will use	Will not use
Total (n = 12,538)	90.1%	9.9%
Customers using the Aqua Line frequently (n = 1,784)	94.5%	5.5%

*5 Drive Plaza: A Web site that publishes information useful for driving holidays, mainly for expressways. Provides search for routes and tolls as well as information regarding tolls, discounts, service areas, and the areas under the jurisdiction of NEXCO East.

*6 Drive Traffic: A Web site publishing traffic information for expressways throughout Japan. Includes mainly real-time traffic restrictions and congestion, congestion forecasts, and scheduled restrictions.

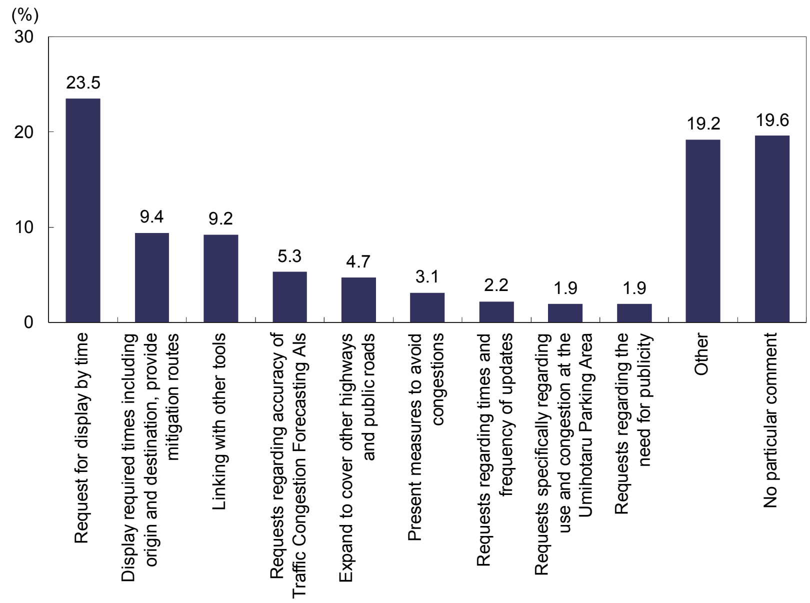

Figure 5　Opinions and requests regarding Traffic Congestion Forecasting AI (compiled from open comment field)

Drive Plaza Web site and began a new trial providing this information in December 2018 [4].

5. Conclusion

This article described Traffic Congestion Forecasting AI, a system that is able to predict traffic congestion in the *future* from Real-time Population Statistics, which estimates *current* population distributions for all of Japan based on mobile telephone network operations data. The article also described trials of Traffic Congestion Forecasting AI conducted in collaboration with NEXCO East on the Tokyo Bay Aqua Line expressway.

We are continuing trials of Traffic Congestion Forecasting AI to verify its effects and any issues and will improve and extend the system based on the results of the trials. We will continue technical development to realize more comfortable driving environments that will enable more drivers to avoid congestion.

REFERENCES

[1] H. Tanaka: "Goodbye Congestion," Nikkei Computer, pp.44–51, Aug. 2018 (In Japanese).
[2] Mobile Spatial Statistics Web page (In Japanese). https://mobaku.jp/
[3] NTT DOCOMO Press Release: "(Notice) NEXCO East and NTT DOCOMO begin trials of 'Traffic Congestion Forecasting AI' on the Tokyo Bay Aqua Line," Nov. 2017 (In Japanese).
[4] NTT DOCOMO Press Release: "'Traffic Congestion Forecasting AI' on the Tokyo Bay Aqua Line provides transit times every 30 minutes," Dec. 2018 (In Japanese).