Technology Reports
 Demand Prediction
 Mobile Kukan Toukei (Demographic Stats from Mobile Phones)
 Deep Learning

 Special Articles on AI Technologies Contributing to Industry and Society

# AI Taxi —Taxi Passenger Demand Prediction Technology for Optimizing Traffic—

Service Innovation Department Satoshi Kawasaki Shin Ishiguro Yusuke Fukazawa IoT Business Department Masanori Fujita Ryohei Suzuki Akihito Makishima

To increase taxi driver revenue, it is important for them to find customers efficiently, even in unfamiliar areas and time periods, and to maximize the time that the taxi is occupied. To address this, NTT DOCOMO has developed technology to predict demand for taxis by area, based on past taxi operation data and statistical information on locations of people. The technology is being offered commercially as a service. The technology enables taxi drivers to check changes in demand in real time, increasing productivity through efficient operation, and also could help reduce taxi customers wait times.

## 1. Introduction

To improve revenue in taxi operation, it is very important to reduce the amount of time taxis are vacant, and maximize the time that they are occupied. Taxi drivers can hope to increase their revenue if they are able to find customers efficiently and increase their occupancy rate, even in unfamiliar areas and time periods and in environments that change constantly, with opening of new roadways and commercial facilities, and conditions that change suddenly, such as when a train is delayed. Improvements in efficiency will also improve productivity, and promote a more work-friendly environment.

In response to these conditions, NTT DOCOMO has devised a method for predicting taxi demand in any given area using data on past taxi rides and localized population statistics<sup>\*</sup>.

<sup>©2018</sup> NTT DOCOMO. INC.

Copies of articles may be reproduced only for personal, noncommercial use, provided that the name NTT DOCOMO Technical Journal, the name(s) of the author(s), the title and date of the article appear in the copies.

<sup>\*</sup> The population statistics used by AI Taxi indicate collective numbers of people for a given area or other attribute, and contain no information that can be used to identify an individual. As such, there is no way for anyone to learn anything about anyone else's movements from these population statistics. The population statistics used by AI Taxi conform to the Mobile Kukan Toukei guidelines indicated below. https://www.nttdocomo.co.jp/corporate/disclosure/mobile\_spatial\_ statistics/guideline/index.html

We also developed a demand-prediction system incorporating the method, and conducted practical testing to verify its utility in cooperation with a taxi operator. The tests demonstrated effects of improving revenue and occupancy rates in all time periods of the trial. We have begun offering this system commercially as our AI Taxi<sup>®\*1</sup> [1] product. By providing demand prediction system results to taxi drivers, taxis can be dispatched to areas where demand is high, as shown in **Figure 1**. This results in reduced wait times for customers, as shown in Figure 2 (1), and can also lead to increased revenue for the taxi business, as shown in Fig. 2 (2) - (4).

# 2. Taxi Passenger Demand Prediction Technology

#### 2.1 Overview

The technology takes input data from Mobile







#### Figure 2 Desired effects from AI Taxi

\_\_\_\_\_

\*1 AI Taxi<sup>®</sup>: A registered trademark of NTT DOCOMO, Inc.

Kukan Toukei (Demographic Stats from Mobile Phones) Real-time Version<sup>\*2</sup>, weather forecasts, past taxi ride data and other sources, and predicts taxi demand on a 500 m mesh<sup>\*3</sup> for 30 minutes in the future (demand is the number of rides to be taken within a given section of the mesh).

To predict future time-series data from past records, generally either an Auto Regressive (AR) model<sup>\*4</sup>, or a further development of AR called an Auto Regressive Integrated Moving Average (ARIMA) model<sup>\*5</sup> is used. These time-series predictive models could also be used for taxi demand, but taxi demand is also affected by other factors, such as weather forecasts and fluctuations in the local population, so we have used a multivariate auto-regressive model<sup>\*6</sup> that incorporates this additional data to improve accuracy. However, with auto-regressive models, including multivariate auto-regressive models, suitable weightings (parameters) of the feature values for the model equations at each mesh point are generally decided by a person. Expanding as a service over wider areas would increase the number of mesh points, so a method for determining such weightings mechanically was needed. As such, we used deep learning techniques to obtain combinations of data and parameters that can improve accuracy from such heterogeneous data. The commercial service utilizes the result from the multivariate auto-regression model or deep learning, whichever yields the more accurate result, as shown in **Figure 3**.

Deep learning is a machine learning<sup>\*7</sup> technique that requires much processing power for the learning process, but is attracting attention recently because increases in hardware performance have made it possible to implement the technique at lower cost than before. Deep learning has already demonstrated dramatically better performance than other machine learning methods in various fields such as image recognition, natural language processing,



Figure 3 Taxi demand prediction technology architecture

- \*2 Mobile Kukan Toukei, Real-time Version: Population statistics generated using the NTT DOCOMO mobile telephone network, able to estimate recent population values on a 500 m grid every ten minutes.
- \*3 Mesh: A grid dividing the country into square sections, along lines of latitude and longitude.
- \*4 AR model: A regressive model, in which result values (objective variables) are modeled using an equation expressed in

terms of affecting values (explanatory variables), but one in which past values of objective variables are used as explanatory variable.

\*5 ARIMA model: A regressive model combining Auto Regressive, Moving Average, and Integrated models. The moving average model sums the differences between past predicted values and actual values, and the integrated model uses the differences for modeling.

and speech recognition. Deep learning uses a neural network<sup>\*8</sup> with many layers (generally four or more) and is able to extract important elements from data, expressing structure and relationships within the data as high-level feature values from very simple feature value inputs, without the need to extract or select advanced feature values or otherwise design a model beforehand.

#### 2.2 Using the Real-time Version of Mobile Kukan Toukei

A major feature of this technology is that it uses the real-time version of Mobile Kukan Toukei to predict demand. Mobile Kukan Toukei<sup>\*9</sup> has been used earlier with past population statistics [2], but to improve the accuracy of taxi demand predictions it is important to use population statistic closer to the current time, so we have used the real-time version of Mobile Kukan Toukei, which can provide such recent data.

As shown in Figure 4, taxi occupancy rates are

higher at times and in locations where population is concentrated more than usual, so use of population data could help in finding latent demand.

There are correlations between population fluctuations and fluctuations in demand for taxis, as shown in Figure 5. At mesh point A, as the population increases, demand for taxis also increases. It is likely that this mesh point includes other transit facilities, such as a train station, and demand increases (high values in the graph) due to people coming from other areas and transferring to taxis. Conversely, at mesh point B, increases in demand for taxis occur roughly five hours after increases in population. It is likely that there is a commercial establishment or event being held at this mesh point, and visitors to the facility spend some time there, and then take a taxi to some other area. This illustrates how changes in population as input data could be used to predict future changes in demand for taxis. However, depending on the characteristics of each area, the delay between



Figure 4 Taxi customer acquisition rate

- \*6 Multivariate auto-regressive model: An auto-regressive model extended for multiple variables. Also called a vector auto-regressive model.
- \*7 Machine learning: A framework that enables a computer to learn useful judgment standards through statistical processing from sample data.
- \*8 Neural network: A type of model that imitates characteristics of the human brain using computers. They use numerical models

that can adjust the degree of connection between nodes using training data, similar to how repeated training changes the strength of connections between neurons in the brain. For machine learning in the past, a designer would have to find characteristics in the input data and design feature values, but a big difference with neural networks is that the training data is given to the neural network, and the feature values are obtained mechanically, similar to what happens in the human brain.



Figure 5 Relationship between changes in taxi demand and population

changes in population and demand for taxis can be different, so to make predictions, we need to determine how much emphasis should be placed at various points earlier in the data for each area. To handle such complex interrelations within the data, we decided to use deep learning and its ability to extract characteristics mechanically.

### 2.3 Demand Prediction Using Deep Learning

Various network structures have been investigated for deep learning, but we have used the Stacked denoising Autoencoders (SdA) model [3], which develops the autoencoder [4] idea further, in our method.

An autoencoder is a network structure that reduces the input data to intermediate data with fewer elements than the input, and then restores it to the same number of elements as the input. It consists of three layers: the input layer representing the input data, a hidden layer representing the intermediate data, and an output layer representing the restored data. The hidden layer has fewer elements than the input data, so when restoring the data, the weightings for less important elements have been reduced, and the weightings

<sup>\*9</sup> Mobile Kukan Toukei: A registered trademark of NTT DOCOMO, Inc.

for elements that have particular importance in reproducing the input data get increased. The network structure used for our method stacks multiple autoencoders, as shown on the right side of Figure 6, enabling it to extract feature values more efficiently and achieve higher accuracy. Important features are extracted through several hidden layers. Also, to handle demand prediction as a regression problem<sup>\*10</sup>, the last autoencoder hidden layer, which extracts the most important feature values, is connected to an output regressive prediction layer<sup>\*11</sup>. In this layer, the values are multiplied by the learned weightings, and the results are added together. Then the final demand prediction results are output. In addition to past taxi ride and population time-series data, statistical

data, such as average rides for the same day of the week and time of the day, is also added to the input layer, as shown on the left in Fig. 6. Note that in the training process with an SdA model, relations between temporal and spatial information are learned mechanically from large amounts of training data, in the form of the weightings on links between each node.

Adding noise to the autoencoder input during training also contributes noise-cancelling (denoising) characteristics. The autoencoders are intended to reconstruct the input data, but by adding faults (or noise) to the input data, and training them to restore the original data which has no faults (or noise), we can expect the effects of faults or noise in the original data to be minimized, and the information



Figure 6 Network structure for predicting taxi demand

\*10 Regression problem: A problem that can be handled as a regression, deriving the objective variable using with an expression in terms of the affecting variables.
\*11 Prediction Layer: A layer that estimates what will appear next

from a given input.

most important for recovering the original data to be extracted preferentially.

## 3. Testing

We conducted tests predicting taxi demand in Tokyo using this technology. We computed how many taxis were expected to find customers in each area on a 500 m mesh in the following 30 minutes. We used data from April 1, 2015 to August 31, 2016 as training data, and from September 1 to September 14, 2016 as evaluation data. Other details regarding the data are given in **Table 1**. Tablets next to the driver's seat displayed a screen like that shown in **Figure 7**, giving the taxi demand prediction results for areas on a map delineated by red lines and enabling drivers to move to areas where predicted demand was higher.

During the testing period, there was some variation in accuracy of predictions by the technology depending on area, but results ranged from 93 to 95%, confirming that the predictions were very accurate. The actual occupancy rates and predicted results are compared in **Figure 8**. Compared with

Table 1	Data	used in	verification	testing
---------	------	---------	--------------	---------

Data sour	ce	Metropolitan Tokyo	
Time perio	bd	Apr. 1, 2015 - Sep. 14, 2016	
Tovi doto	No. of cars	4.400	
	Acquisition frequency	Every 5 to 10 s	
Dopulation data	Resolution	500 m	
Population data	Acquisition frequency	Every 10 min	
Deinfall data	Resolution	250 m	
naimali data	Acquisition frequency	Every 10 min	



Figure 7 Screen displaying demand prediction results



Figure 8 Comparison of demand prediction results and actual rides

results one month before the tests, sales amounts for drivers using the demand prediction system increased by an average of 1,400 yen per driver per day in all times tested, which was more than the average increase for all drivers. The taxi occupancy rates, which are an index of efficiency in getting customers, also increased.

### 4. Conclusion

This article described a technology used in the AI Taxi Service that predicts future demand for taxis. Matching supply with demand is important for achieving efficient operation. The technology enables the supply of unoccupied taxis to be optimized by predicting demand ahead of time. In future work, we will also study dispatch control mechanisms that consider global optimization.

#### REFERENCES

- [1] AI Taxi Service Web site: https://www.nttdocomo.co.jp/biz/service/aitaxi/
- [2] Okajima et al.: ""Mobile Spatial Statistics" Supporting Development of Society and Industry - Population Estimation Technology Using Mobile Network Statistical Data and Applications -," NTT DOCOMO Technical Journal, Vol.14, No.3, pp.4–50, Jan. 2013.
- [3] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio and P. A. Manzagol: "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," Journal of Machine Learning Research, pp.3371–3408, Dec. 2010.
- [4] G. E. Hinton and R. R. Salakhutdinov: "Reducing the dimensionality of data with neural networks," Science, Vol.313, Issue 5786, pp.504–507, 2006.