Marked Temporal Point Processes for Trip Demand Prediction in Bike Sharing Systems

Maya OKAWA^{†a)}, Yusuke TANAKA[†], Takeshi KURASHIMA[†], Hiroyuki TODA[†], *and* Tomohiro YAMADA[†], *Nonmembers*

SUMMARY With the acceptance of social sharing, public bike sharing services have become popular worldwide. One of the most important tasks in operating a bike sharing system is managing the bike supply at each station to avoid either running out of bicycles or docks to park them. This requires the system operator to redistribute bicycles from overcrowded stations to under-supplied ones. Trip demand prediction plays a crucial role in improving redistribution strategies. Predicting trip demand is a highly challenging problem because it is influenced by multiple levels of factors, both environmental and individual, e.g., weather and user characteristics. Although several existing studies successfully address either of them in isolation, no framework exists that can consider all factors simultaneously. This paper starts by analyzing trip data from real-world bike-sharing systems. The analysis reveals the interplay of the multiple levels of the factors. Based on the analysis results, we develop a novel form of the point process; it jointly incorporates multiple levels of factors to predict trip demand, i.e., predicting the pick-up and drop-off levels in the future and when overdemand is likely to occur. Our extensive experiments on real-world bike sharing systems demonstrate the superiority of our trip demand prediction method over five existing methods.

key words: point process, bike-sharing system, trip prediction, human mobility

1. Introduction

With the acceptance of shared social systems, public bike sharing services have become popular worldwide [1], [2]. The services consist of multiple docking stations (stations hereafter) spread across the city where bicycles are available for short-term rental. A user can pick up a bicycle at any station (pick-up), use it for short-distance trip, and drop off it to any other station with an open dock (drop-off). Such systems benefit a city in many ways [2]; they offer an affordable mobility option for the citizens; alleviate traffic congestion; reduce exhaust gas emission; and promote public health. Bike sharing systems have been developed in many major cities including New York, Chicago and Tokyo, and most of them have been growing in size since launch.

The most important task in operating a bike sharing system is managing the bicycle supply in each station; avoiding station over-demand, i.e., being either empty or full. Stations with high pick-up demand may run out of bicycles, while ones with higher drop-off demand tend to fill up with bicycles and run out of docks to park them.

Manuscript publicized June 17, 2019.

a) E-mail: maya.ookawa.af@hco.ntt.co.jp

Over-demand is a critical problem for both users and system operators. To avoid it, the system operator must move bicycles from overcrowded stations to under-supplied ones. The most common strategy is to redistribute the bicycles by truck or to incentivize users to return their bicycles to the under-supplied stations. In either case, trip demand prediction plays a crucial role in scheduling proper redistribution in terms of number and timing; if we can accurately predict the pick-up and drop-off demand at each station and when over-demand will occur, we can redistribute the proper number of bicycles to the right place at the right time in order to prevent over-demand.

Predicting trip demand is a highly challenging problem because bicycling behavior is influenced by multiple levels of factors, both environmental and individual [3]. Environmental factors include weather [4]–[6] and social events [7]. Previous studies [5], [6] present aggregate-level models based on clustering-based regression to capture the environmental factors. Another line of work [8], [9] explores the impact of individual factors on trip demand. Individual factors are defined as any factor associated with each individual trip, including user characteristics and type of trip. For example, Zhang et al. [9] propose an individuallevel model that considers user type and attributes to predict each user's trip behavior. However, the different levels of factors have, up to now, been explored separately, no framework exists that integrates the multiple levels of key factors simultaneously.

We develop a novel method that incorporates both environmental and individual factors for predicting trip demand, where the goal is to predict the pick-up and drop-off demand in the future and the timing of the over-demand. We formulate the problem of trip prediction in the framework of the marked temporal point process. The marked temporal point process is an effective mathematical framework for modeling the occurrence of events as marks over time [10]; it estimates an intensity function that describes the occurrence rate of events in continuous time periods. Point process based methods have been widely used in many disciplines and fields including seismology and finance. Recent works explore the application of the point process method to mobility modeling, e.g., POI (points of interest) recommendation [11], [12], trip purpose detection [13]. But some challenges remain in applying the point process methods to the goal of bike trip modeling. First, they cannot incorporate the environmental factors such as weather. Second, the

Manuscript received November 5, 2018.

Manuscript revised April 8, 2019.

[†]The authors are with NTT Service Evolution Laboratories, Yokosuka-shi, 239–0847 Japan.

DOI: 10.1587/transinf.2018OFP0004

functional form of the intensity must be carefully designed by researchers so that it appropriately captures the reality. Little is known about the interplay between the environmental and individual factors and their impact on trip demand.

In this paper, we first empirically investigate how trip demand is jointly influenced by the multiple levels of factors, and then, based on the empirical insights, develop a novel functional form of point process intensity. In our framework, trips are organized as sequences of pick-up events, each of which is associated with individual factors. We introduce a set of intensity functions with covariates of environmental factors to model these event sequences, where individual factors are treated as marks. This formulation allows us to effectively capture both environmental and individual factors.

Finally, we conduct experiments on the data of three real-world bike-sharing systems and show that the proposed method consistently outperforms existing methods with regard to the trip demand prediction task.

The main contributions of this paper are as follows:

- Empirical analysis: We examine how the multiple levels of factors jointly influence trip demand using realworld bike sharing datasets (Sect. 3).
- Model building: We construct a novel method for trip demand prediction based on the analysis results. The proposed method enables more accurate prediction by taking the multiple levels of influential factors into account (Sect. 5).
- Evaluations: We conduct extensive experiments on the data of real-world bike sharing systems from three cities. The proposed method achieves better predictive performance than all existing methods on all datasets (Sect. 6).

2. Related Work

In this section, we first review the literature on bike-sharing systems and then position our contributions. With the rapid development of bike-sharing systems, large amounts of bike trip data are being generated. Several services publish trip datasets for public use. Many studies that use these datasets with different aims have been presented, e.g., station place-ment [14]–[16], rebalancing [6], [17], [18] and trip demand prediction. This paper focuses on trip demand prediction, which is one of the most important issues in operating bike-sharing systems.

The literature on trip demand prediction continues to expand. They can be broadly divided into two categories: aggregate-level methods and individual-level methods. Aggregate-level methods focus on predicting the aggregated number of trips. They first discretize time into bins of fixed width (e.g., 1-hour) and estimate the number of bicycles that will be picked up and dropped off in each time bin. Time-series models, such as ARMA [19] and ARIMA [20], were introduced to predict the number of available bicycles for each station a few minutes/hours ahead. In a recently paper, Froehlich et al. [21] use a Bayesian network to predict the number of bicycles for each station, while Schuijbroek et al. [18] employ a Markov chain to predict station availability. Chen et al. [7] developed a dynamic cluster-based method for over-demand prediction, which exploits periodic patterns and environmental factors such as the weather. Yang et al. [5] built a random forest model (RF) which includes periodic features (i.e., hour of day and day of week) and the weather (e.g., temperature and visibility) as covariates. Liu et al. [6] proposed Meteorology Similarity Weighed K-Nearest-Neighbor (MSWK) regression, which considers the similarity of related features, i.e., hour of day, day of week and the weather. However, all these aggregatelevel methods still have their own limitations. First, they fail to incorporate influential factors associated with individual trips, i.e., user identifier, types of trip (e.g., round-trip vs. one-way), membership (e.g., subscriber or temporary user) and user profile (e.g., age, gender, etc.) into their models. Second, they divide time into equal sized bins. Bin size is a critical model parameter and should be determined ad hoc (i.e. they are not continuous-time models). Furthermore, they are designed to predict the number of trips; they cannot be used to predict the timing of over-demand. The successful prediction of over-demand timing is essential in allowing bike-sharing system operators to decide when is the optimal time to move bicycles as well as the optimal number to move. If they know the timing of over-demand, they can move the proper number of bicycles at the right time.

Individual-level methods such as point process models [22], [23] can handle individual trips and offer the possibility of overcoming the limitations, as they view pick-up and drop-off times as discrete events occurring in continuous time. A major drawback of this approach is that it does not consider environmental factors such as the weather. Point process models have been successfully applied in similar domains such as mobility modeling [11], [12] and social activities [24]–[27]. However, to the best of our knowledge, no published method can handle multiple levels of the factors simultaneously.

This paper uses the framework of a marked temporal point process to propose a novel method that covers the multiple levels of factors including environmental context and individual factors.

3. Analysis of Bike Trips

In this section, we provide a series of empirical observations about the interplay between individual and environmental factors, on which our method is founded.

3.1 Data

For explanation, we use a collection of trips generated from bike-sharing systems, from Sep. 2015 to Aug. 2016. For each trip, these records provide a pick-up station, the datetime of pick-up, the drop-off station, duration and user type. The user type information includes membership (i.e., temporary user or subscriber), gender and age. This analysis focuses on membership which has a significant impact on trip demand [9]. Further, we have weather data which includes temperature, humidity, wind speed and visibility distance. The datasets are described in detail in Sect. 6. We report the findings from just the New York City dataset due to the space limitation, the other datasets show similar results.

3.2 Empirical Observations

In this section we examine how multiple levels of factors affect bike pick-up demand (Sect. 3.2.1) and trip duration (Sect. 3.3.2). The key observations are listed in each subsection.

3.2.1 Pick-Up Demand

Previous studies showed that pick-up demand changes in the environmental factors such as weather [5], [6], [28] as well as individual factors such as user type [8], [9]. Here we further analyze *how* these two levels of factors influence the pick-up demand *collectively*.

By examining the temporal evolution of bike pick-up demand, we find that the weather variables have non-linear effects on trip demand, which can be approximated by third degree polynomials (**P1**). Figure 1 shows the average number of pick-up events per day for subscribers and temporary users across different weather variables: temperature (left) and humidity (right). The dotted lines represent the results of third degree polynomial fitting. This result also shows the *collective* influence of the weather and user type factors. The demand variations under different weather conditions are quite different for the different user types (**P2**). It is apparent that temporary users are more sensitive to weather stress than subscribers.







Fig. 2 Trip durations for different temperatures for different user types

3.2.2 Trip Duration

Most existing works [4], [5] simply fit a single distribution to trip durations between each pair of stations, without considering the influential factors. This paper investigates in more detail whether and how trip duration changes under different environmental and individual contexts.

Our analysis indicates that the variation in trip duration is mostly explained by user type (P3). Figure 2 plots the log-normal distribution best-fit to the normalized histogram of trip durations for different temperatures. Figure 2(a) is the distribution of trip duration for all users; Fig. 2 (b) is the subscriber distribution. According to Fig. 2(a), trip durations seem to fluctuate with the weather. Average trip duration is longer with nice temperatures than with cool or hot temperatures. This can be explained by different trip behavior across different user type (registered subscribers or temporary users). As mentioned in the previous subsection, temporary users use the bikes much less than subscribers in bad weather conditions such as too hot or cold temperatures. Also, the average trip durations vary significantly between subscribers and temporary users [9]. We confirm that the same discussion holds for the other weather variables and periodic factors. This knowledge allows us to simplify the model (see Sect. 5).

4. Notations and Definitions

In this section, we first introduce some notations and definitions and then formally define the trip demand prediction problem. The notations used in this paper is summarized in Table 1.

A bike trip treated as the tuple (u, v, t, τ, m) which means a user of type *m* picked up a bicycle from station *u* at time *t* and dropped off the bicycle at station *v* at time τ . The duration of each trip is defined as $\Delta = \tau - t$. Formally, let $S = \{s_1, \ldots, s_K\}$ be a set of stations, where *K* denotes the number of stations. We have a set of bike trip histories $\mathcal{D} = \{(u_i, v_i, t_i, \tau, m_i)\}_{i=1}^n$ with a total of *n* instances. Each element in \mathcal{D} is an event consisting of pick-up station, dropoff station, pickup time, drop-off time, and user type; where user of type $m_i \in \mathcal{M}$ picks up a bicycle from station $u_i \in S$ at continuous time $t_i \in \mathbb{R}^+$, and drops off the bicycle at station $v_i \in S$ at continuous time $\tau_i \in \mathbb{R}^+$ ($t_i < \tau_i$). Trip duration is represented by $\Delta_i = \tau_i - t_i$. Figure 1 illustrates an example

Table 1Important notations.

Symbol	Description
S	Set of stations, $S = \{s_1, \ldots, s_K\}$
Κ	Number of stations
u_i/v_i	Pick-up/drop-off station of trip i
t_i/τ_i	Pick-up/drop-off time of trip i
Δ_i	Trip duration of trip i , $\tau_i - t_i$
\mathcal{M}	Set of user types
m_i	User type of trip <i>i</i>
W_t	Weather vector at time t



Fig. 3 Example of bike trip data: The x-axis presents a timeline of events for each station.

of bike trip data. In this figure, the black arrows denote trips between stations. Each black arrow is a pick-up event and each red diamond marker is a drop-off event. At the bottom of Fig. 4, we plot the number of pick-up events (black line), the number of drop-off events (red line) and the number of bicycles (yellow line) at station s_K .

We further have a set of weather data $\mathcal{W} = \{W_t\}_{t \in \mathbb{R}^+}$. $W_t = (F_t, H_t, S_t, V_t, R_t)$ is a weather vector corresponding to time *t*, where F_t , H_t , S_t , V_t and R_t denote temperature, humidity, wind speed, visibility distance, and rainfall, respectively.

The trip demand prediction task tackled in this paper is defined as:

Problem 1: Given bike trip data \mathcal{D} and weather data \mathcal{W} , we aim to solve the following two sub-tasks:

- Pick-up and drop-off number prediction: What number of bicycles will be picked up or dropped off at each station in time window [*t*_{start}, *t*_{end}]?
- Over-demand time prediction: What is the expected time at which over-demand will occur at each station?

5. Trip Demand Prediction

In this section, we first provide the necessary background on marked temporal point processes. We then introduce our model and derive an optimization algorithm for our method. Finally, we provide the prediction procedure.

5.1 Marked Temporal Point Process

Marked temporal point process is a random process of event occurrences over time. Let (t, v) be an event, represented by the pair of time $t \in \mathbb{R}^+$ and another marker information v (e.g., station). Marked temporal point process is governed by non-negative intensity function $\lambda(t, v)$. Intensity $\lambda(t, v)$ represents the probability of the occurrence of a new instance of v in small time window [t, t + dt]. The functional form of the intensity function is designed to capture the phenomena of interest [29]. Typically, intensity $\lambda(t, v)$



Fig. 4 Framework of the proposed method

reduces to $\lambda(t, v) = \lambda(t)f(v|t)$, where f(v|t) is the conditional mark density function and $\lambda(t)$ represents the probability of an event occurring in small time window [t, t + dt]. Given historical observations $\mathcal{H} = \{(t_i, v_i)\}_{i=1}^{N_H}$ with a total of N_H events, the likelihood function in observation time window [0, T] is given by

$$p(\mathcal{H}|\lambda(t,v)) = \left[\prod_{(t_i,v_i)\in\mathcal{H}}\lambda(t_i)f(v_i|t_i)\right]\exp\left(-\int_0^T\lambda(t)dt\right).$$
(1)

5.2 Proposed Method

This paper aims to design a trip demand prediction method that properly incorporates both environmental and individual factors. Our method follows a few basic assumptions drawn from previous studies: (A1) pick-up and drop-off demand varies with circadian and weekly patterns [5], [6], [28]; and (A2) trip duration follows a log-normal distribution [4]. Additionally, we apply the insights gained from empirical observations (P1)-(P3) to our method.

Figure 4 shows the framework of our proposed method. From the left, we have two data sources, i.e., bike trip data and weather data. The two kinds of data are processed by two models: 1) *pick-up model* models pick-up demand at each station by a marked temporal point process, where drop-off stations and individual factors (e.g., user type) are represented as marks; 2) *trip duration model* describes trip durations between each pair of stations with log-normal models (A1). Finally, our method predicts drop-off demand by combining the pick-up demand and trip duration. The following subsection details the models of our method.

5.2.1 Pick-Up Model

The pick-up model is built upon the marked temporal point process. Here we explain how to incorporate the aforementioned assumptions into the point process framework. We assume that the set of pick-up events at station *u* follows a marked point process with intensity function $\lambda_u(t, v|W_t)$ with covariates of environmental factors (weather variables) W_t as follows:

$$\lambda_u(t, v | W_t) = f_u(v | t) \lambda_u(t | W_t), \tag{2}$$

where $\lambda_u(t, v|W_t)$ indicates the occurrence probability that a user picks up a bicycle at time *t* and station *u*, and $f_u(v|t)$

is the conditional probability that the user who picks up the bicycle at time t and station u will drop it off at station v. This formulation enables us to incorporate the environmental factors.

We design $f_u(v|t)$ based on assumption (A1): drop-off demand changes with circadian and weekly patterns. First we introduce a relative time from the beginning of the week $d \in [0, 6]$ days for each time t. We then use the Gaussian Naive Bayes approach to model the popularity $f_u(v|d)$ of station v at time d [30]. In particular, we rewrite $f_u(v|t)$ as $f_u(v|t) \propto p_u(t|v = z)p_u(z)$ using Bayes' theorem, where dropoff station v is regarded as class $z \in S$. We fit a Gaussian distribution to $p_u(t|v = z)$, and a multinominal distribution to $p_u(z)$.

We model $\lambda_u(t|W_t)$ as the product of two factors: temporal factors and weather factor $\alpha_{u,m}(W_t)$, as in

$$\lambda_u(t|W_t) = \sum_{m \in \mathcal{M}} \alpha_{u,m}(W_t) \lambda_{u,m}(t).$$
(3)

Here we introduce the individual functions for each type of user m based on empirical insight (P3), different user types have quite different trip demand variations.

Temporal factor: To describe a periodic pattern with one-week cycle (A1), we define the temporal factor $\lambda_{u,m}(t)$ by using a mixture of *J* periodic Gaussian kernels [31]:

$$\lambda_{u,m}(t) = \sum_{j=1}^{J} x_j^{u,m} \exp\left(-\sin^2\left(\pi \frac{t-\omega_j}{l}\right)/h^2\right),\tag{4}$$

where ω_j is the fixed center of the *j*-th Gaussian kernel, h is the standard deviation, and $x_i^{u,m}$ is the weight for the *j*-th kernel of user type m at station u. l is the periodicity hyper-parameter that represents the distance between the repetitions. For example, by setting the periodicity hyperparameter value l = 24 hours, we can enforce a circadian repetition. In this study, we use l = 168 (i.e., 7×24) hours to capture a weekly repetition. J specifies the approximation of the periodic pattern. Here we choose J = 168, which represents the weekly periodic pattern with a granularity of 1-hour. The top row of Fig. 7 in the experimental section depicts the temporal factor $\lambda_{u,m}(t)$ learned from a real-world data. We can see that the mixture-based formulation provides a flexible model, allowing to capture the complex pattern with a weekly periodicity. In the experiment, we set h = 60 minutes and $\omega_i = j \times 60$ minutes.

Weather factor: In order to model the weather influence (P1), we adopt the following multivariate polynomial regression model:

$$\alpha_{u,m}(W_t) = \exp\left(b_0^{u,m} + \sum_{k=1}^K \mathbf{b}_k^{u,m} W_t^{\circ k}\right)$$
(5)

where W_t is the five-dimensional weather vector, $\mathbf{b}_k^{u,m}$ and $b_0^{u,m}$ are the five-dimensional regression vector and the regression intercept for user type *m* and station *u*, $A^{\circ k}$ is the *k*-fold Hadamard product (Hadamard power) of *A*. We use

an exponential function to ensure non-negativity of the intensity. $\alpha_{u,m}(W_t)$ represents the non-linear influence of the weather on trip demand. We choose K = 2 (i.e., the binomial regression) in this paper, and it is sufficient to describe the weather influence as seen in the experimental section.

Note that this can be easily generalized to high-degree multinomials. We denote the parameter set of $\lambda_u(t|W_t)$ as $\Psi = \{\{x_i^{u,m}\}_{i=1}^J, \gamma, \mathbf{b}_1^{u,m}, \dots, \mathbf{b}_K^{u,m}, b_0^{u,m}\}.$

5.2.2 Trip Duration Model

Trip duration model follows the basic assumptions: it follows a log-normal distribution (A2) and mainly depends on user type m (P3). We assume that trip duration Δ between each pair of stations (station u and v) and each type of user m is given by the following log-normal distribution:

$$p_{u,v}^{(m)}(\Delta) = \frac{1}{\Delta \sigma_{u,v}^{(m)} \sqrt{2\pi}} \exp\left(-\frac{(\log \Delta - \mu_{u,v}^{(m)})^2}{2\sigma_{u,v}^{(m)^2}}\right),\tag{6}$$

where $\{\mu_{u,v}^{(m)}, \sigma_{u,v}^{(m)}\}_{m \in \mathcal{M}, u \in \mathcal{S}, v \in \mathcal{S}}$ are the model parameters.

5.3 Learning

The log-likelihood of the pick-up model can be written by

$$\mathcal{L} = \sum_{i=1}^{n} \log \lambda_{u_i}(t_i | W_{t_i}) - \sum_{u \in S} \int_0^T \lambda_u(t | W_t) dt$$
$$+ \sum_{i=1}^{n} \log f_{u_i}(v_i | t_i).$$
(7)

We note that this reduces to the independent optimization problem for $\lambda_u(t|W_t)$ and $f_u(t, v)$. For $\lambda_u(t|W_t)$, because the objective function is differentiable with respect to all the parameters, Ψ , gradient-based optimization algorithms can be used. We leverage the stochastic gradient decent (SGD) algorithm to solve the optimization problem. For $f_u(t, v)$, following the general procedure described in [30], we obtain the closed form solution of $\{\theta_u^z, \phi_u^z\}_{z \in S}$. For our trip duration model, the simple formulation (Eq. (6)) allows us to derive a closed form solution for all parameters. Due to the page limit, we omit details of the inference procedures.

5.4 Prediction

Given the learned intensity function $\lambda_u(t, v|W_t)$ and the distribution of trip durations $p_{u,v}^{(m)}(\Delta)$, we can predict the trip demand for each station, *u*. Here, we provide the prediction procedures for the two sub-tasks, i.e., pick-up and drop-off number prediction and over-demand prediction.

5.4.1 Pick-Up and Drop-Off Number Prediction

For pick-up number prediction, we directly integrate the intensity function: $\int_{t_{\text{start}}}^{t_{\text{end}}} \lambda_u(t) dt$, which gives the expected number of pick-up events in time window [$t_{\text{start}}, t_{\text{end}}$]. The

drop-off intensity of station v is given by

$$\lambda'_{v}(t|W_{t}) = \sum_{u \in \mathcal{S}} \sum_{m \in \mathcal{M}} \int \alpha_{u,m}(W_{t}) \lambda_{u,m}(t) p_{u,v}^{(m)}(t-t') dt'.$$
(8)

For drop-off number prediction, we derive the expected number of drop-off events at station v by integrating the drop-off intensity: $\int_{t_{start}}^{t_{end}} \lambda'_v(t|W_t) dt$. As this integration is analytically intractable due to the complexity of $\lambda'_v(t|W_t)$, we perform numerical integration.

5.4.2 Over-Demand Time Prediction

The over-demand time, \hat{t}_o , at station *u* is given by

$$\hat{t}_{o} = \min\left\{t_{o}: N_{u}(t_{c}) + \int_{t_{c}}^{t_{o}} [\lambda'_{u}(t|W_{t}) - \lambda_{u}(t|W_{t})]dt < 1\right\},$$
(9)

where t_c is the current time and $N_u(t_c)$ is the current number of bicycles (or docks) at station u. As t_o is analytically intractable from Eq. (9), the solution is numerically searched for by a scan over $t_o \in [t_c, t_f]$ in stepwise fashion, where t_f is the end time of the test period.

6. Experiment

In this section, we evaluate the predictive performance of our model on real-world datasets with regard to predicting pick-up/drop-off number and when over-demand will occur.

6.1 Datasets

We used datasets from three cities: New York City, Washington, D.C. and San Francisco. Each dataset includes two sub-datasets: bike trip data and weather data. For San Francisco, bike trip data also contains station status history. All the bike trip data are publicly available. The weather data can be downloaded via the weather underground API[†]. The details are as follow.

- New York City (NY)^{††}: The trip data is taken from the bike-sharing system in New York City for the period Sept 2015 to Aug 2016 (The observation period is 1 year). Each trip record contains five fields: pick-up and drop-off dates/times, pick-up and drop-off stations, trip duration and type of user (i.e., subscriber or temporary user).
- *Washington D.C.* (*D.C.*)^{†††}: For Washington D.C., we obtained data from the bike-sharing system for the period Jan 2015 to Sept 2016. Each dataset has the same format as the NY dataset.
- San Francisco $(SF)^{\dagger\dagger\dagger\dagger\dagger}$: The dataset was collected in

Table 2Basic dataset statistics.					
Data sources		NY	D.C.	SF	
Bike trip data	# Stations # Trips (Subscriber) # Station status	633 12,906,339 11,367,000 –	402 3,252,709 2,581,413 -	74 313,685 280,091 per minute	
Meteo- rological data	Temperature (°C) Humidity (%) Visibility (km) Wind Speed (mph) Rain fall (mm)	[-18, 35] [0, 100] [0, 137] [0, 16] [0, 19]	[-5, 36] [0, 100] [0, 48] [0, 16] [0, 9]	[0, 33] [0, 100] [0, 390] [0, 16] [0, 11]	
Observation period		9/1/2015 - 8/31/2016			

San Francisco for the period Sept 2015 to Aug 2016. For each trip, the dataset provides pick-up and drop-off dates/times, pick-up and drop-off stations and trip duration. SF dataset also contains station status history, i.e., the number of bicycles and docks available in stations recorded almost every minute.

It should be noted that prediction performance might depend on test period. Therefore, we split each dataset into 4 chronological subsets each of three month duration (e.g., Jan 2015 - Mar 2015, Apr 2015 - Jun 2015 etc.) and performed independent runs for each subset. In order to remove the bias of the day of week, we selected the last 7 days of each subset as the test set and used the remaining data as training set. For over-demand prediction, we ran the step-wise search algorithm described in Sect. 5.5 (Eq. (11)) to every day in the test period.

Table 2 provides the statistics of the datasets.

6.2 Evaluation Metrics

For each subtask, we introduce different metrics. For pickup and drop-off number prediction, we discretize time into 1-hour bins and compare the number of pick-up and dropoff events at each station in each time bin with ground truth by using mean absolute error (MAE):

$$MAE = \frac{\sum_{n=1}^{N} \sum_{r=1}^{R} |\hat{y}_{n,r} - y_{n,r}|}{RN},$$
(10)

where *N* denotes the number of time steps in the test period and *R* denotes the number of stations. $\hat{y}_{n,r}$ is the predicted number of pick-ups or drop-offs at the *r*-th station for the *n*-th time step; $y_{n,r}$ is the ground truth. For the overdemand time prediction task, we introduce penalized root mean square error (PRMSE). Given predicted over-demand times $(\hat{t}_r^1, \ldots, \hat{t}_r^M)$ and actual over-demand times (t_r^1, \ldots, t_r^N) at *r*-th station, PRMSE is defined as the square root of the following equation:

$$\sum_{r=1}^{R} \left(\frac{1}{Q} \sum_{i=1}^{Q} (\hat{t}_{r}^{i} - t_{r}^{i})^{2} + \mathbb{I}[M > N] \sum_{i=N+1}^{M} (T - \hat{t}_{r}^{i})^{2} + \mathbb{I}[N > M] \sum_{i=M+1}^{N} (T - t_{r}^{i})^{2} \right),$$
(11)

where $Q = \min(M, N)$. $\mathbb{I}[\cdot]$ is an indicator function, which

[†]https://www.wunderground.com/weather/api/

^{††}https:// www.citibikenyc.com/system-data

^{†††}http://www.capitalbikeshare.com/system-data

^{††††}https://www.fordgobike.com/system-data

indicates 1 when the condition holds, and 0 otherwise. PRMSE measures absolute difference between the predicted times and the actual times at each station. The second and third terms respectively penalize excessive and insufficient number of times predicted by our method [32].

6.3 Baseline Methods

For pick-up and drop-off number prediction, we compare the proposed model with the following five baseline methods.

- *Historical Mean (HM)* [21]: Estimates the pick-up (drop-off) number during the specific time period as averaged pick-up (drop-off) number in historical data.
- Auto-Regressive and Moving Average (ARMA) [33]: Makes one-step ahead prediction of the pick-up (dropoff) number based on the historical data by exploiting its temporal correlation.
- Inhomogeneous Poisson Process Inference (IPPI) [22]: Constructs individual Poisson process models for pickup and drop-off time. These models do not include the correlation between pick-up and drop-off across stations, and does not consider trends.

Note that the above three methods do not consider any external factors such as weather and user types as they are originally based on the historical data and cannot be easily extended to capture the external influences.

- *Multi-similarity-weighted k-NN (MSWK)* [6]: Outputs the aggregated number of pick-ups and drop-offs for the given future time period and weather variables, where the similarity of the input values is taken into consideration.
- *Random-forest-based method (RF)* [5]: Predict the aggregated number of pick-ups and drop-offs based on time factors (i.e. time of day and day of week) and weather variables by using Random forest model.

Both MSWK and RF, which are intended to model the aggregated data, do not incorporate influential factors associated with individual trips (e.g., user type). Hence we only consider environmental factors (e.g., weather) and time factors (e.g., day of week) for these methods in the experiment. Also, all the existing methods, except for IPPI, are focus on predicting the aggregated number of trips; thus not to applicable for over-demand time prediction. For this task, we compare the proposed method and IPPI.

6.4 Evaluation Results

6.4.1 Pick-Up and Drop-Off Number Prediction

Figure 5 shows the predictions output by the six different methods for the three datasets, i.e., NY, D.C. and SF, where 5(a) presents the MAE for pick-up prediction and 5(b) is the MAE for drop-off prediction. In all cases, the proposed



Fig.5 Mean absolute error (MAE) in predicting (a) pick-up and (b) drop-off number. Lower is better. Comparison to baselines.

method outperforms the baseline methods and the differences are significant (two-sided t-test: p < 0.01). This result suggests the proposed method effectively incorporates both environmental factors (weather) and individual factors (user type). On the whole, ARMA performs worse than the other methods, demonstrating the benefit of considering periodicity. For pick-up number prediction, the clusteringbased methods, RF and MSWK, yield comparable results with HM, even though they consider additional factors such as the weather. RF provides slightly less accurate predictions than HM on two datasets (D.C. and SF). This could be due to the limited amount of data (see Table 2); for the NY dataset, RF performs better. Assuming a parametric form of the underlying process, the proposed method is robust against data sparsity. For drop-off number prediction, the proposed method outperforms the existing methods. This suggests the appropriateness of the trip duration model. We obtained similar results using other metric (e.g. Mean Averaged Percentage Error).

6.4.2 Over-Demand Time Prediction

To evaluate the effectiveness of the proposed method for over-demand time prediction, we use the learned intensity function of IPPI (following the same procedure for overdemand time predictions as in the proposed method). Because only the SF dataset contains station status history, we conduct the experiment on the SF dataset. We used the timing at which the number of bicycles or docks turns to be zero as the ground truth for the actual over-demand time. Table 3 gives the predictions output by IPPI and the proposed method for the SF dataset. Clearly, the proposed method performs significantly better than IPPI. According to Table 3, the proposed model results in about 13% smaller MAE than the baseline method. This is because the proposed method captures both environmental and

Table 3Mean absolute error (MAE) for over-demand prediction on theSF dataset. The numbers in parentheses indicate the standard errors.



Fig. 6 Comparison of the estimated over-demand time by IPPI (middle) and proposed method (right) with actual time (left) around San Francisco Ferry Building on a typical weekday (Feb. 24, 2016).



Fig.7 Learned intensity function. Top row: temporal factor; bottom row: weather factor. block.

individual factors, while IPPI considers only the periodic pattern. Figure 6 further supports this, which compares the estimated over-demand time with the actual time around San Francisco Ferry Building on a typical weekday, Feb. 24, 2016. The left plot is the actual time, the middle plot is the prediction results of IPPI, and the right plot corresponds to the values output by the proposed method. It can be seen that the proposed method provides more reasonable predictions than IPPI.

6.4.3 Case Study

Figure 7 presents the learned intensity at the most popular station in Washington D.C. from Aug 20 to Aug 27 2016, where the top row shows the temporal factor; and the bottom row is the weather factor. In this figure, the pink line shows the intensity functions of temporary users and the black line shows those of subscribers. According to the top row, the registered subscribers mainly rent bicycles on the weekdays. Their trip demand has two peaks around 9:00 and 18:00 on the weekdays. This shows that the temporary users tend to use bicycles during the daytime. In the bottom row of Fig. 7, we can see that the trip demand of temporary users changes more dramatically with the weather than that of subscribers, as confirmed in the preliminary analysis (Sect. 3.2.1). These results demonstrate that the proposed method yields

valuable insights about how each type of users responds to environmental changes. These insights can enhance bicycle redistribution e.g., making incentives more effective.

7. Conclusion and Future Work

In this paper, we studied the problem of trip demand prediction focusing on the sub-problems of predicting trip demand and over-demand times. We first gathered and analyzed data from three cites: NY, D.C. and SF. Our empirical analysis of the real-world bike-sharing datasets demonstrated that trip demand is jointly influenced by multiple factors and that modeling them simultaneously is critical to achieving accurate predictions. Based on the insights, we used the marked temporal point process framework to construct a novel method that jointly captures the multiple levels of factors including the individual factors such as user type and environmental factors such as weather. We conducted experiments on the data of three real-world bikesharing systems. Our results demonstrated the superiority of the proposed method for trip demand prediction over five existing methods. We note that the proposed method is directly applicable to various other domains, including taxi dispatch and targeted ad. In future work, we will perform experiments on a wider variety of datasets including taxis, car-sharing and public transit to demonstrate the superiority of the proposed method in various application domains.

References

- P. DeMaio, "Bike-sharing: History, impacts, models of provision, and future," Journal of Public Transportation, vol.12, no.4, pp.41–56, 2009.
- [2] S. Susan, G. Stacey, and Z. Hua, "Bike sharing in europe, the americas, and asia," Transportation Research Record, vol.1, no.2, pp.159–167, 2010.
- [3] S.L. Handy and Y. Xing, "Factors correlated with bicycle commuting: A study in six small us cities," International Journal of Sustainable Transportation, vol.5, no.2, pp.91–110, 2011.
- [4] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic prediction in a bike sharing system," SIGSPATIAL, 2015.
- [5] Z. Yang, J. Hu, Y. Shu, P. Cheng, J. Chen, and T. Moscibroda, "Mobility modeling and prediction in bike-sharing systems," MobiSys 2016, pp.165–178, 2016.
- [6] J. Liu, L. Sun, W. Chen, and H. Xiong, "Rebalancing bike sharing systems: A multi-source data smart optimization," KDD, pp.1005–1014, 2016.
- [7] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.M.T. Nguyen, and J. Jakubowicz, "Dynamic cluster-based over-demand prediction in bike sharing systems," Ubicomp, pp.841–852, 2016.
- [8] X. Zhou, "Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in chicago," Cartographica, vol.10, no.10, e0137922, 2015.
- [9] J. Zhang, X. Pan, M. Li, and P.S. Yu, "Bicycle-sharing system analysis and trip prediction," Mobile Data Management, vol.2016-July, pp.174–179, 2016.
- [10] M. Jacobsen, Point process theory and applications: marked point and piecewise deterministic processes, Springer Science & Business Media, 2006.
- [11] H. Kim, T. Iwata, Y. Fujiwara, and N. Ueda, "Read the silence: Well-timed recommendation via admixture marked point processes,"

AAAI, pp.132-139, 2017.

- [12] N. Du, H. Dai, R. Trivedi, U. Upadhyay, M. Gomez-Rodriguez, and L. Song, "Recurrent marked temporal point processes: Embedding event history to vector," KDD, pp.1555–1564, 2016.
- [13] P. Wang, "Human mobility synchronization and trip purpose detection with mixture of hawkes processes," KDD, pp.495–503, 2017.
- [14] L. Chen, D. Zhang, G. Pan, X. Ma, D. Yang, K. Kushlev, W. Zhang, and S. Li, "Bike sharing station placement leveraging heterogeneous urban open data," Ubicomp, pp.571–575, 2015.
- [15] J. Zhang, X. Pan, M. Li, and P.S. Yu, "Bicycle-sharing systems expansion: station re-deployment through crowd planning," SIGSPATIAL, p.2, 2016.
- [16] J. Liu, "Functional zone based hierarchical demand prediction for bike system expansion," KDD, pp.957–966, 2017.
- [17] C. Médard de Chardon, G. Caruso, and I. Thomas, "Bike-share rebalancing strategies, patterns, and purpose," Journal of Transport Geography, vol.55, pp.22–39, 2016.
- [18] J. Schuijbroek, R.C. Hampshire, and W.-J. van Hoeve, "Inventory rebalancing and vehicle routing in bike sharing systems," European Journal of Operational Research, vol.257, no.3, pp.992–1004, 2017.
- [19] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system," Pervasive and Mobile Computing, vol.6, no.4, pp.455–466, 2010.
- [20] J.W. Yoon, F. Pinelli, and F. Calabrese, "Cityride: A predictive bike sharing journey advisor," Mobile Data Management, pp.306–311, 2012.
- [21] J. Froehlich, J. Neumann, and N. Oliver, "Sensing and predicting the pulse of the city through shared bicycling," IJCAI, no.2, pp.1420– 1426, 2009.
- [22] R. Alvarez-valdes, J.M. Belenguer, E. Benavent, J.D. Bermudez, E. Vercher, and F. Verdejo, "Optimizing the level of service quality of a bike-sharing system," Omega, 2015.
- [23] N. Gast, "Probabilistic forecasts of bike-sharing systems for journey planning," CIKM, pp.703–712, 2015.
- [24] N. Navaroli and P. Smyth, "Modeling response time in digital human communication," AAAI, pp.278–287, 2013.
- [25] S.H. Yang and H. Zha, "Mixture of mutually exciting processes for viral diffusion," ICML, pp.1–9, 2013.
- [26] A. Zarezade, A. Khodadadi, M. Farajtabar, H.R. Rabiee, L. Song, and H. Zha, "Correlated cascades: Compete or cooperate," AAAI, pp.238–244, 2015.
- [27] Q. Zhao, M.A. Erdogdu, H.Y. He, A. Rajaraman, and J. Leskovec, "Seismic: A self-exciting point process model for predicting tweet popularity," KDD, pp.1513–1522, 2015.
- [28] A. Singla, M. Santoni, G. Bartok, P. Mukerji, M. Meenen, and A. Krause, "Incentivizing users for balancing bike sharing systems," AAAI, no.2014, pp.723–729, 2015.
- [29] O. Aalen, O. Borgan, and H. Gjessing, Survival and event history analysis: a process point of view, Springer Science & Business Media, 2008.
- [30] I. Rish *et al.*, "An empirical study of the naive bayes classifier," IJCAI workshop on empirical methods in artificial intelligence, pp.41–46, 2001.
- [31] D.J. MacKay, "Introduction to gaussian processes," NATO ASI Series F Computer and Systems Sciences, vol.168, pp.133–166, 1998.
- [32] M. Lukasik, P.K. Srijith, T. Cohn, and K. Bontcheva, "Modeling tweet arrival times using log-gaussian cox processes," EMNLP, no.Sept., pp.250–255, 2015.
- [33] J.W. Böse, H. Hu, C. Jahn, X. Shi, R. Stahlbock, and S. Voß, Computational Logistics: Second International Conference, ICCL Springer, 2011.





Maya Okawa Researcher, NTT Service Evolution Laboratories, NTT corporation. She received the B.S. degree in physics from Sophia University in 2012 and the M.S. degree in particle physics from University of Tokyo in 2014. She joined NTT in 2014. She is currently pursuing the Ph.D. degree in computer science at the Department of Science and Technology Intelligence, Graduate School of Informatics, Kyoto University, Japan. Her research interests include machine learning and data mining.

Yusuke Tanaka Researcher, NTT Service Evolution Laboratories, NTT corporation. He received the B.E. degree in engineering from Kobe University in 2011 and the M.Inf. degree in informatics from Kyoto University in 2013. He joined NTT in 2013. He is currently pursuing the Ph.D. degree at the Department of Systems Science, Graduate School of Informatics, Kyoto University, Japan. His research interests include machine learning and data mining. He is a member of the Information Processing Society

of Japan (IPSJ) and the Japanese Society for Artificial Intelligence (JSAI).





Takeshi Kurashima received the B.S. degree from Doshisha University, Japan, in 2004, and the M.S. and Ph.D. degrees in Informatics from Kyoto University, Japan, in 2006 and 2014, respectively. He is currently a Research Engineer at NTT Corporation, Japan. He was a visiting scholar at Stanford University from 2016 to 2017. His current research interests include data mining, machine learning, and recommender systems. He is a member of IEICE.

Hiroyuki Toda Senior Research Engineer, Supervisor, NTT Service Evolution Laboratories, NTT corporation. He received a B.E, and M.E. in materials science from Nagoya University in 1997 and 1999, and a Ph.D. in computer science from University of Tsukuba in 2007. He joined NTT in 1999. His current research interests include information retrieval, and data mining. He is a member of the Information Processing Society of Japan (IPSJ), the Japanese Society for Artificial Intelligence (JSAI), the

Database Society of Japan (DBSJ), and the Association for Computing Machinery (ACM).



Tomohiro Yamada Project manager & Executive Research Engineer, Supervisor, Proactive Navigation Project, NTT Service Evolution Laboratories. He received the M.S. in electronics engineering from Niigata University in 1992 and the MBA from the University of Birmingham, UK in 2007. He joined NTT in 1992, where he has been engaged in research on content distribution systems and technologies to support the human-robot communication. He is Director of i-RooBO Network Forum Institute,

Chair of IEICE Technical Committee on Life Intelligence and Office Information Systems, and also a member of the Association for Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), and the Information Processing Society of Japan (IPSJ).