Towards bicycle demand prediction of large-scale bicycle sharing system

Yufei HAN[*] Latifa Oukhellou[†] Etienne Come[‡]

Submitted For Publication
93rd Annual Meeting of the Transportation Research Board
January 12, 2014

Word Count:

Number of figures:	6 (250 words each)
Number of tables:	1 (250 words each)
Number of words in texts:	4821
Total:	4821 + 7*250 = 6571 (MAX 7500)

[*]Corresponding Author, IFSTTAR/GRETTIA, France, yfhan.hust@gmail.com
[†]IFSTTAR, latifa.oukhellou@ifsttar.fr
[‡]IFSTTAR, etienne.come@ifsttar.fr

Abstract

We focus on predicting demands of bicycle usage in Velib system of Paris, which is a large-scale bicycle sharing service covering the whole Paris and its near suburbs. In this system, bicycle demand of each station usually correlates with historical Velib usage records at both spatial and temporal scale. The spatio-temporal correlation acts as an important factor affecting bicycle demands in the system. Thus it is a necessary information source for predicting bicycle demand of each station accurately. To investigate the spatio-temporal correlation pattern and integrate it into prediction, we propose a spatio-temporal network filtering process to achieve the prediction goal. The linkage structure of the network encodes the underlying correlation information. We utilize a sparsity regularized negative binomial regression based variable selection method to learn the network structure automatically from the Velib usage data, which is designed to highlight important spatio-temporal correlation between historical bicycle usage records and the bicycle demands of each station. Once we identify the network structure,a prediction model fit well with our goal is obtained directly. To verify the validity of the proposed method, we test it on a a large-scale record set of Velib usage.

1 Introduction

Urban shared-mobility has attracted more and more attentions for both academic researchers and city policy-makers to build livable and sustainable communities. Bicycle sharing systems (BSS) have been very successfully deployed in many metropolitans in the world. The main motivation is to provide users with free or rental bicycles especially suited for short-distance trips in urban areas, thus reduces traffic congestion, air pollution and noise that leads to high economical and social cost. In Europe, BSSs are most popular in southern European countries. Thanks to their unquestionable success [7, 4], more and more European cities works to provide this mode of mobility in order to modernize the city planning. In France, the first implementation of BSS was in Lyon in 2005 (it is called Velib'v). Nowadays,BSSs have been launched in twenty French cities, including Paris, one of the most large-scale BSSs in France (it is called Velib).

The fundamental issue of BSS study is to understand bicycles mobility patterns and regulate availability of the bicycles in the urban network. Due to differences between city blocks in social activities and functions, demands of short-distance trips usually form non-uniform spatio-temporal patterns. Some BSS stations tend to face large demand of bicycles during specific time periods. The planning department of BSS thus needs to forecast the bicycle demand variations at each station, in order to balance the bicycle loading in the system. Several studies [10, 2, 16, 13] have shown the usefulness of analyzing the data collected by BSSs operators. The redistribution of bicycles can benefit from the analysis of statistical bicycle usage patterns [8, 14, 1].

Fruitful progress as these studies have achieved, there is still an open question of BSS service demand prediction. What are the factors affecting bicycle demands of one specific station ? How do the bicycle demands of one station correlate with historical bicycle usage patterns ? Previous works attacked the prediction of BSS demands mainly from two aspects. In [11], the spatio-temporal bicycle usage patterns of the whole network are extracted from BSS data using clustering algorithms and historical average of BSS demands corresponding to each pattern is employed to achieve the forecast goal. On the other hand, [11] ignores the spatio-temporal correlation and construct station-wise forecast only depending on the temporal dynamics of each station. However,both of them don't investigate the spatio-temporal factors of bicycle demands with respect to individual stations explicitly.

The contribution of this paper aims at solving the investigation of the spatio-temporal bicycle usage patterns through a network structure learning procedure. Our objective is to exhibit the important factors affecting the Velib service demands at each station in order to provide a direct solution to the open problem. Based on the analysis, we are able to achieve our goal of Velib service demand forecast for the whole network immediately.

This paper is organized as follows. In Section.2, a general description of Velib system is given. Section.3 describes the proposed analysis methodology performed on the collected Velib usage data. Section.4 is devoted to illustrate the identified spatio-temporal factors on Velib service demand at each station. Section.5 presents the capabilities of the proposed method in forecasting Velib service demands, in comparison with two other baseline regression technologies utilizing no explicit spatio-temporal factors of bicycle demands at all. Section.6 concludes the whole paper.

2 Velib system description

Velib is designed to facilitate sightseeing and public transportation in urban areas of Paris in 2007. Total 7,000 bicycles are initially distributed to 750 fixed stations. In 2008, Velib system is further extended to 20,000 bicycles over 1,208 fixed service stations. The system serves 110,000 short-distance trips on average per day. Most Velib stations are located within Paris. A small proportion of stations are distributed in the near suburbs in order to extend sharing service.

Velib provides a non-stop 24-hour bicycle rental service. Each Velib station has an automatic rental payment terminal and 8 to 70 bicycle docking positions. There are totally 40,000 docking positions available in the system. Each bike is locked to its docking position electrically. Users can either purchase a short-term (daily or weekly) usage of the bicycle or charge a annual pass card. To fetch the bicycle, the user need to show his usage card to RFID terminals equipped with each docking position in order to unlock the bicycle. The first 30 minutes for short-term rental and the first 45 minutes for annual rental of every trip is free of charge. Users can return the bicycles easily to any station at any time.

The Velib usage dataset used in our work is composed of over 2,500,000 trip records during five months (April, June, September, October, December) in 2011. To obtain stable usage patterns, we remove trips with time duration of less than one minute and with the same station as origin and destination from the data set. These fake trips correspond to users' mis-operations. Finally we reserve trips from 1188 stations in the city. Based on the trip data, we count the number of bicycles departing from and entering each station per hour. The number of departing bicycles per hour at each station represents the hourly profile of Velib service demand at this station. Figure. 1 shows the histogram of time duration of each bicycle trip. The y axis represents the number of the trips with different levels of time duration length. Most trips are finished within less than 2 hours, which is consistent with the fact that Velib system serves the short-distance mobility in the city. We count the average number of departing bicycles per hour for all the stations to evaluate activity level of Velib usage globally in the system. Figure. 2 shows the difference in hourly Velib usage between weekdays and weekends. A cyclostationarity pattern can be seen in Velib usage during weekdays. Three peaks of weekday usage can be observed in Figure 2: the most significant two correspond to the public commutes (from 8am to 10am and from 6pm to 9pm), while the third one from 11 am to 13 pm corresponds to the lunch break. They represent travel patterns of public transportation, such as home-office and office-restaurant patterns. In contrast, the morning peak usage disappears during the weekends. Velib usage gradually reaches the maximum in the afternoon, reflecting the travel patterns of the leisure time.

These statistics give a general profile of Velib usage patterns in the system. In this paper, based on the extracted Velib usage count data, we aim to achieve short-term forecast of the bicycle demands at each station by analyzing inter-station spatio-temporal correlation of bicycle usages.

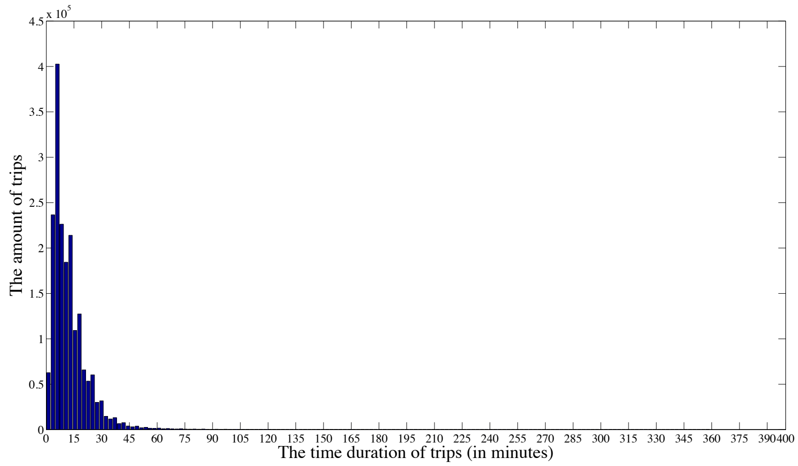


Figure 1: Statistics of time duration of each trip

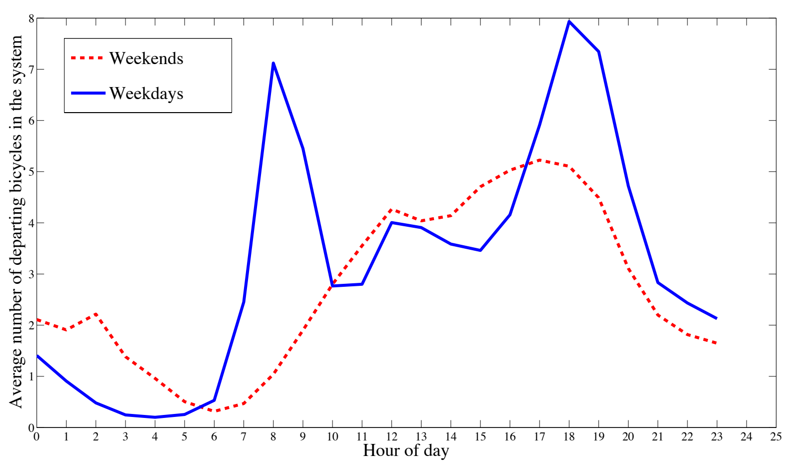


Figure 2: Average number of departing per hour during weekdays (continuous blue line) and weekends (dashed red line).

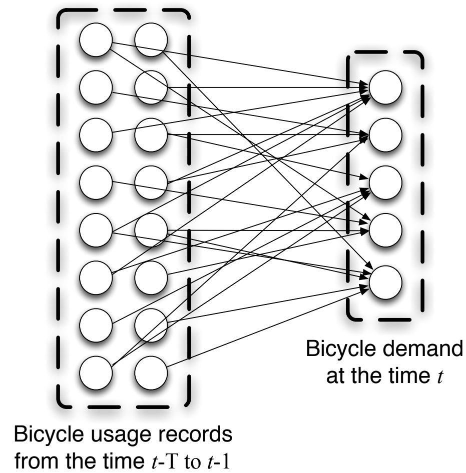


Figure 3: Spatio-temporal predictive network

3 Velib demand prediction through learning of spatio-temporal network structure

We use $X_{i,t}^{out}$ and $X_{i,t}^{in}$ $i = 1, 2, 3..., n$ to note the number of bicycles departing from and entering the station i at the time t respectively. $n = 1188$ is the number of Velib stations. Each pair of $X_{i,t}^{out}$ and $X_{i,t}^{in}$ indicate the temporal dependent usage pattern of Velib service of each station. Notably $X_{i,t}^{out}$ represents bicycle demands of the station i at the time t, which is the target of the predictive analysis. In our work, we assume that the temporal dynamics of bicycle demand is a stationary markov process. It means the $X_{i,t}^{out}$ only depends on the recent Velib usage records from $t - T$ to $t - 1$. Considering most Velib trips are short-distance travels that last no more than 2 hours, the markovian assumption is reasonable and we set T to be 2. With this setting, the conditional dependence between historical Velib usage records and the bicycle demands to be predicted in the system can be described using a spatio-temporal network $G = \{\Delta_1, \Delta_2, E\}$,as in Figure. 3. Δ_1 and Δ_2 are two sets of nodes. Each node of Δ_1 is the historical Velib usage $X_{i,j}^{out}$ and $X_{i,j}^{in}$ in the system from the time $j = t - T$ to $j = t - 1$. Nodes of Δ_2 correspond to bicycle demands at the time t $X_{i,t}^{out}$ for each station, which are the prediction target. E is the set of directed edges linking nodes of Δ_1 and Δ_2. The linkage structure of the network represents spatio-temporal correlations between the historical Velib usage records $X_{i,j}^{out}$ and $X_{i,j}^{in}$ $(i = 1, 2, 3..., n, j = t - T, t - T + 1, ..., t - 1)$ and the prediction target $X_{i,t}^{out}$ $(i = 1, 2, 3..., n)$.

A link from one node in Δ_1 to another in Δ_2 represents the existence of spatio-temporal correlation that is useful for prediction between the two nodes On the contrary, if two nodes from different sets are separated with no linkage, they are irrelevant with respect to the prediction goal. The interior linkage between nodes within Δ_1 and Δ_2 are ignored since we aim to construct a prediction model instead of a generative model to simulate spatio-temporal dynamics of Velib usage.

For each node s in Δ_2, we note the $N(s)$ is the set of nodes in Δ_1 that are linked to the node s. In the domain of network structured data analysis, $N(s)$ is also defined as a neighbor set of s in the network G. Identifying neighborhood structure for each node in Δ_2 finally achieves to reconstruct the network structure. In our work,neighborhood selection is the key step to analyze the useful spatio-temporal correlation between Δ_1 and Δ_2 . In statistics,it is an intuitive solution to calculate covariances between nodes of Δ_1 and Δ_2 and judge the correlation level given the covariances. However, in large-scale urban area, the number of nodes in Δ_1 (n times T) is usually much larger than the volume of the available historical Velib usage records. The derived covariances easily generates fake correlation between nodes [9]. In machine learning research, estimating correlation structure of high-dimensional data is a popular topic. Most solutions are proposed by performing global sparsity constraint on covariance matrix of data to find the strong correlation patterns [9]. This kind of methods assumes that all nodes follow a joint normal distribution. It fits the joint data distribution with a generative gaussian random field [9], which is beyond the goal of the prediction task. Furthermore, the joint normal distribution assumption is not valid for count data. The alternative solutions investigate the correlation of each node in the network with the others by performing sparsity-inducing regression. Treating the concerned node s in Δ_2 as the regression target and the other nodes in Δ_1 as covariate of regression, the sparsity-inducing regression generates sparse regression coefficients. Zero coefficients indicate conditional independence between the corresponding nodes of Δ_1 and the concerned node s, while coefficients with distinctively large magnitudes indicate strong correlation between them. Benefited from the regression-oriented neighborhood selection scheme, the identified correlation structure is intrinsically selected to suit the prediction task. Furthermore,we can obtain a prediction model immediately after fixing the network structure. Therefore, we address the issue of network structure learning following this idea.

3.1 $L1$-norm regularized negative binomial regression for neighborhood selection

For each node $X_{\Delta_{2,s}}$ in Δ_2, we identify its neighbors in Δ_1 with a sparsity regularized regression model. The mathematical expression is defined as follows:

$$\begin{aligned}\left(\theta_k^s\right) &= min_{\theta^s} f\left(X_{\Delta_{2,s}}, X_{\Delta_1,k}, \theta^s\right) + \lambda\|\theta^s\| \\ k &= 1,2,...2*n*T\end{aligned} \tag{1}$$

In Eq 1,$X_{\Delta_1,k}$ $k = 1,2,...2*n*T$ represents all nodes in Δ_1, representing historical records of the number of bicycles entering or departing from each station within the time frame from $t-T$ to $t-1$. θ^s is the regression coefficient vector of the same dimension as $X_{\Delta_1,k}$ $k = 1,2,...2*n*T$ for the station s. Each component θ_k^s is the regression coefficient for the corresponding node $X_{\Delta_1,k}$. $\|\theta\|$ is the $L1$ norm of θ^s, defined as sum of absolute values of each θ_k^s. The function $f\left(X_{\Delta_{2,s}}, X_{\Delta_1}, \theta^s\right)$ is a generalized linear regression model to suit different types of the regression target.λ is the penalty parameter balancing the sparsity-introducing regularization and the regression cost. In [5, 15, 18], f is a least regression function on $X_{\Delta_{2,s}}$ suited for continuous variable. [17] defines f as a logistic regression function since the regression target is a binary in Ising model. In our work, $X_{\Delta_{2,s}}$ represent the number of bicycle rented at the station s, which is integer count data. Therefore,we define f as negative binomial regression model as in Eq 2.

$$
\begin{aligned}
f &= \log P_{nb}\left(X_{\Delta_{2,s}}\right) \\
P_{nb}\left(X_{\Delta_{2,s}}\right) &= \frac{\Gamma\left(X_{\Delta_{2,s}}+\psi\right)}{\Gamma\left(X_{\Delta_{2,s}}+1\right)\Gamma\left(\psi\right)}\left(\frac{\psi}{\psi+\mu}\right)^{\psi}\left(\frac{\mu}{\mu+\psi}\right)^{X_{\Delta_{2,s}}}
\end{aligned}
\tag{2}
$$

where $\mu = e^{\sum_k \theta_k^s X_{\Delta_1,k}}$ is the mean of the negative binomial distribution $P_{nb}\left(X_{\Delta_{2,s}}\right)$. In generalized linear model, it is fitted using the exponential link function based on the covariate $X_{\Delta_1,k}$. ψ is the dispersion parameter of negative binomial distribution. Γ is the Gamma function defined as $\Gamma\left(n\right) = (n-1)!$. Negative binomial regression [12] is a generalized poisson regression to fit the dispersed count data with the larger variance than the mean. The standard poisson regression is formulated as follows:

$$
P_{poisson}\left(X_{\Delta_{2,s}}\right) = \frac{e^{-\mu_p}\mu_p}{X_{\Delta_{2,s}}!} \tag{3}
$$

where $\mu_p = e^{\sum_k \theta_k^s X_{\Delta_1,k}+\varepsilon}$,including a random intercept ε as random noise. μ_p is the product of the link function $e^{\sum_k \theta_k^s X_{\Delta_1,k}}$ and the random factor $\nu = e^{\varepsilon}$. The link function represents the non-linear relation between the poisson mean and the input covariate. Negative binomial regression is then derived by performing gamma prior probability on the random factor ν and integrating out ν, as expressed in the followings:

$$
P_{nb}\left(X_{\Delta_{2,s}}\right) = \int_0^{\infty} P_{poisson}\left(X_{\Delta_{2,s}}\right) h\left(\nu\right) d\nu \tag{4}
$$

where $h = \frac{\delta^{\psi}}{\Gamma(\psi)}\nu^{\psi-1}e^{-\nu\delta}$ is the gamma prior probability on ν. δ is the shape parameter of the gamma distribution. In poisson regression,both the mean and variance of poisson distribution equals to μ. However, in the Velib count data, the count data $X_{\Delta_{2,s}}$ are more dispersed. The variance is distinctively larger than the mean. By introducing the gamma prior probability in Eq. 4, the mean and the variance of $X_{\Delta_{2,s}}$ are modeled respectively as μ and $\mu + \frac{\mu^2}{\psi}$ in negative binomial regression. The dispersion parameter ψ enables more flexibilty to fit variance with different magnitudes. When ψ goes into infinity, negative binomial regression is reduced to standard poisson regression.

Performing the $L1$ norm based penalization leads to a sparse θ^s vector. Only a small proportion of components θ_k^s have distinctively large magnitudes, while the others are exactly zeros or have very small magnitudes approaching to zero. Through this way, the nodes $X_{\Delta_1,k}$ with distinctively large non-zero $theta_k^s$ are the most important spatio-temporal factors affecting the bicycle demands $X_{\Delta_{2,s}}$. It indicates the existence of linkage between the corresponding historical Velib usage records and $X_{\Delta_{2,s}}$.The weight of this link is the value of the coefficient θ_k^s The rest extremely weak or zero coefficients have no weight in predicting, thus they indicate no linkage connecting the corresponding nodes in Δ_1 to $X_{\Delta_{2,s}}$. The sparse structure of θ^s gives rise to a compact spatio-temporal correlation structure in the network. Once we fix the network structure according to $theta^s$, it is then direct to construct a negative binomial regression model to predict the bicycle demands of the station s. Given all Velib historical usage information in Δ_1 as input, the prediction for all 1188 stations proceeds as a network filtering procedure, providing estimation of nodes in Δ_2. Discriminating power of this generalized linear model is verfied in previous works [12].The predictive model integrates the spatio-temproal correlation of Velib usage records. The sparse structure of the network linkage improves the model compactness by removing irrelevant Velib usage information. In [18, 17], Xu et al proves that the $L1$ norm regularization gives the asymptotically correct estimation of neighborhood structure if the underlying neighborhood is sparse. In Velib service, the bicycle demand denotes the short-distance trip custom in Paris. The bicycle demands at one station are only correlated with those of a specific group of stations. Therefore, this characteristic implies a underlying sparse spatio-temporal correlation structure. The penalization parameter λ is fixed by cross-validation in the neighborhood selection procedure. It is chosen as the one that minimizes the regression error while reserving the sparsity of θ in cross-validation.

3.2 Solution to regularized negative binomial regression

We assume the training data involves total m days of Velib usage data. Each day contains 24-hour records of $X_{i,t}^{out}$ and $X_{i,t}^{in}$ $i = 1, 2, 3..., n, t = 1, 2, 3, ..., 24$. As a result, we can sample 22 time frames of length 2 to form the set X_{Delta_1} on each day. In this setting, we can construct a training data set containing $22 * m$ pairs of $X^j_{Delta_{1,k}}$ $(k = 1, 2, 3..., 2 * n * T)$ and $X^j_{Delta_{2,s}}$ $(j = 1, 2, 3..., 22 * m)$. The objective function for estimating θ^s is then formulated as follows:

$$\theta^s = min_{\theta^s} \sum_{j=1,2,3...,22*m} f\left(X^j_{\Delta_{2,s}}, X^j_{\Delta_1}, \theta^s\right) + \lambda \|\theta^s\| \tag{5}$$

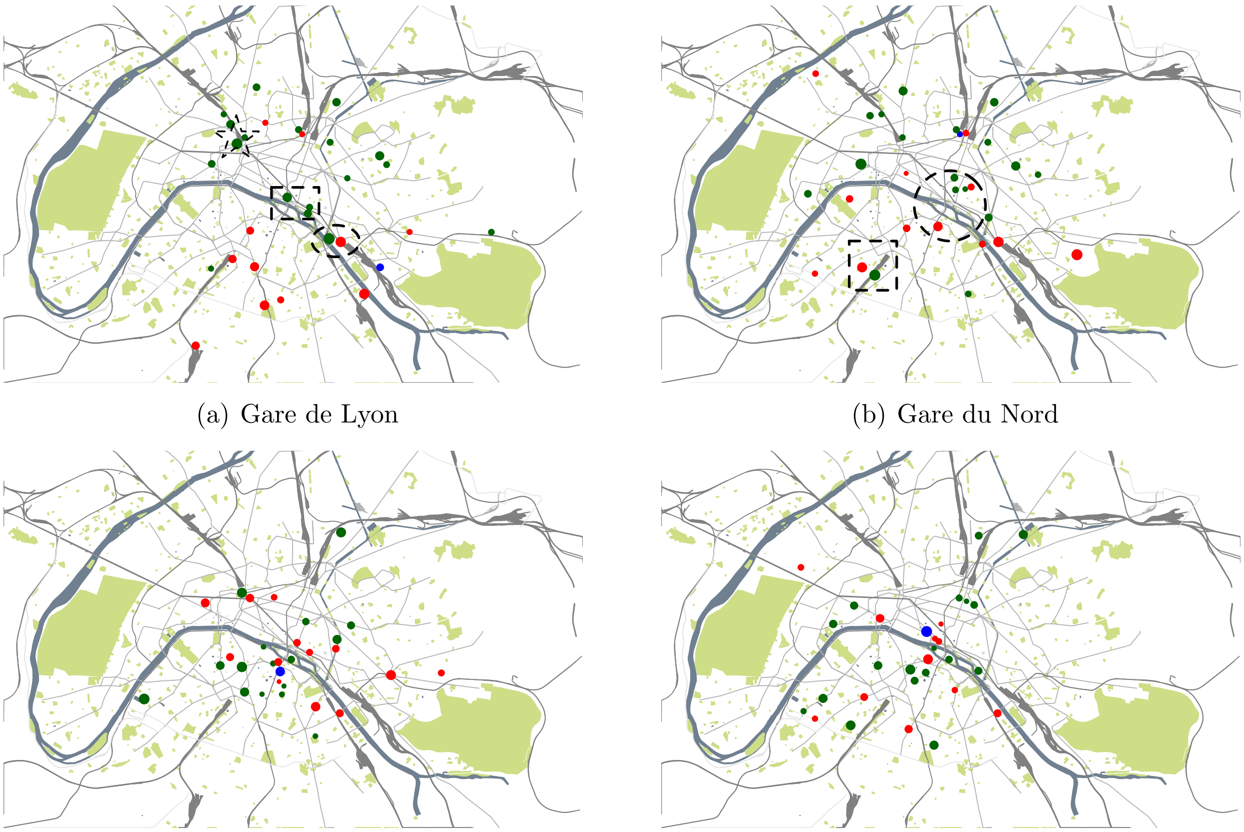
Given a fixed penalization parameter λ, minimizing the cost function in Eq. 5 with f as the negative binomial regression equals to solve a regularized maximum likelihood problem. However,f is not convex due to simultaneous optimization with respect to θ^s and the dispersion parameter ψ. To address problem, an interior point optimization method [3] is applied to relax the original minimization procedure and search for an approximated solution. The basic idea is to optimize the cost function Eq. 5 with respect to only one variable at each time, while the other one is fixed. Each subproblem of optimization is convex and easy to solve by taking the first-order conditions and making them equal to zero. The two first-order optimum conditions, one for θ and one for ψ, are presented as:

$$\sum_{j=1,2,3,...,22*m} \frac{X^j_{\Delta_{2,s}} - \mu_j}{1+\psi^{-1}\mu_j} + sgn(\theta^s) = 0$$
$$\sum_{j=1,2,3,...,22*m} \left(\frac{\log 1+\psi^{-1}\mu_j - \sum_{l=1,2,3,...,X^j_{\Delta_{2,s}}-1}}{\left(\psi^{-1}\right)^2} + \frac{X^j_{\Delta_{2,s}} - \mu_j}{\psi^{-1}\left(1+\psi^{-1}\mu_j\right)} \right) = 0 \tag{6}$$

where $\mu_j = e^{\sum_k \theta^s_k X^j_{\Delta_1,k}}$ and sgn indicates the signs of all θ^s_k in θ^s. Original difficult jointly optimizing problem is relaxed to alternative updates of the parameters θ and ψ iteratively until convergence. Since the non-convexity of the cost function, the alternative update procedure doesn't guarantee to achieve a global optimum. A proper warm-start allows the alternative optimization procedure to achieve reasonably good solution fast. In our work, we follow the idea of [12] to use a poisson regression parameter to initialize θ and ψ in Eq. 4. For each station, the convergence is achieved for each fixed λ after 50 iterations on average.

4 Spatial-temporal correlation structure of Velib usage

The sparse structure of the regression coefficients θ^s indicates the historical Velib usage records $X^{in}_{i,j}$ and $X^{out}_{i,j}$ ($j = t-1, t-2$) that are the most informative for the prediction task. It unveils the compact spatio-temporal correlation pattern with respect to the station s. Magnitudes of the non-zero regression coefficients are proportional to the correlation level of the corresponding historical Velib usage records with the prediction target. In the followings, we illustrate the learned spatio-temporal correlation structures for four stations. Two of them corresponds to Velib stocking points located around the rail-way stations in Paris (Gare du Nord and Gare de Lyon). The other two are located in the down-town area of Paris, near the places of interests (Saint German des Pres and Louvre). Geographical locations of these four stations are carefully selected. Velib usage records of these stations represent typical Velib usage patterns of common home-office travels and sightseeing-oriented travels in Paris. Most short-distance mobility in Paris belong to either of the two travels. For each station s, if the sparse regression coefficient θ^s has less than 20 non-zero entries, we illustrate all the correlated historical Velib usage records corresponding to the non-zero regression coefficients. Otherwise, we select the correlated historical records corresponding to the 20 non-zero regression coefficients of the largest magnitudes. The blue pots in the figures illustrate the concerned station s. The red pots indicate the station i ($i \in 1, 2, 3, ..., 1188$) corresponding to selected $\left\{X^{out}_{i,j}\right\}$ during the precedent 2 hours that are the most correlated with $X^{out}_{s,t}$, while the green pots are the the station i ($i \in 1, 2, 3, ..., 1188$) corresponding to the most correlated $\left\{X^{in}_{i,j}\right\}$ during the precedent 2 hours.



(a) Gare de Lyon

(b) Gare du Nord

(c) Saint Germain des Pres

(d) Louvre

Figure 4: Typical spatio-temporal correlation structure.

In Figure 4(a) and 4(b), we can find both local and non-local spatial-temporl correlation structures. For example in Figure 4(a), Velib service demand near Gare de Lyon is strongly correlated with historical Velib usage information at the entrance of Gare de Lyon (noted by a circular region)and Hotel de Ville (noted by a square region) that locates closely to Gare de Lyon. Besides, Velib usage records around Gare de Saint-Lazare (noted by a star-shaped legend) also present high-level correlation with the prediction target. In Figure 4(b), Velib demand at Gare du Nord has a non-local correlation with Velib usage records at Gare Montparnasse (noted with a circular legend) and the center of Paris (noted with a square legend). The existence of the spatio-temporal correlation doesn't necessarily indicate the existence of physical bicycles flows between the selected stations and the the concerned station s. It means that the historical Velib usage patterns at those stations provides the most critical information in prediction the Velib demand at the station s. This spatio-temporal correlation structure is arisen by either bicycle flow interactions or similar temporal dynamic patterns of Velib usage. The former presents a local correlation at most time and can be investigated further by looking into trip records, while the latter usually presents a non-local correlation strcuture and can not be captured by the trip data. It provides complementary information for prediction of Velib service demand at the concerned station. In Figure 4(c) and 4(d), the spatio-temporal correlation structure are more local than that in the first two figures, Most selected correlated stations locate within the neighborhood of the concerned station. This phenominon represents the characteristics of sightseeing mobility patterns. Different from public transportation, Velib usages for sightseeing are concentrated near the places of interests of the city. In Paris, most places of interests are distributed near Boulevard de Saint-Germain des Pres and Musee du Louvre, along the Seine river. Therefore, originations and destinations of sightseeing travel by Velib are usually within the same area. Besides, temporal variation of bicycle usage for sightseeing is normally inconsistent with that of daily commute. Thus, the bicycle flow interaction becomes the most importation factor affecting the Velib usage patterns of the last two stations, which in turn gives rise to the local spatio-temporal correlation structure.

5 Experiments on spatio-temporal prediction of bicycle usage demand

In this section, we illustrate the prediction performance of the proposed method. In all 152 days of Velib usage count data, we choose each day in turn for testing regression performance and use all the others for learing the network linkage structure θ^s ($s = 1, 2, 3..., 1188$). For the testing day, given Velib usage count of precedent 2 hours ($t-1$ and $t-2$),the task is to predict the bicycle demand $X^{out}_{:,t}$ at the time t for all 1188 stations based on the learned θ^s, where $t = 1, 2, 3, ...22$. The obtained θ^s ($s = 1, 2, 3.., 1188$) are directly used to estimate the mean value μ of the negative binomial regression model in Eq. 2. The mean value μ is the forecast of bicycle demands. The absolute error between the predicted bicycle demands for each station s at each time t is obtained during the testing process. The sample mean and sample variance of the absolute errors are used to evaluate prediction accuacy.

We compare the proposed method with two other prediction schemes. The first one is an intuitive solution. It constructs a poisson regression model for each station independently to achieve temporal prediction,ignoring the spatio-temporal correlation of Velib usage unveiled in the last section. The input of the poisson model for the station s is the Velib usage data $X_{s,j}^{in}$ and $X_{s,j}^{out}$ $(j = t-2, t-1)$. The output is the bicycle demand $X_{s,t}^{out}$ at the time t. The parameters of each poisson regression model are estimated using the quasi-newton optimization. We name it as station-wise poisson regression hereafter.

The other one is designed to make use of the spatio-temporal patterns of Velib usage in an inexplicit way, It combines nearest neighbor regression technologies and state space temporal dynamic model to achieve prediction of the bicycle demand at each station simulataneously [6].In this method, we treat Velib usage count $X_{:,t}^{in}$ and $X_{:,t}^{out}$ for all stations at the time t as a multivariate vector Y_t of $1188 * 2 = 2376$ dimensions, considering both the volume of bicycles enterring and departing from each station. Based on the training set of 151 days, we form a matrix $Y^{\mathbf{train}} \in R^{3624*2376}$ by integrating daily records of all 151 days into the sequence of hourly records of $151 * 24 = 3624$ hours. Each row is defined as Y_t and the rows are arranged following the temporal order of all 151 days in the training set. We then employ Principle Component Analysis (PCA) to decompose the matrix $Y^{\mathbf{train}}$ and project each row $Y_t^{\mathbf{train}}$ to a low-dimensional subspace. We calculate the first k principle eigen vectors of $Y^{\mathbf{train}}$ corresponding to the largest spectrum energy. These principle eigen vectors form a projection matrix $P \in R^{3624*k}$, with the eigen vectors arranged as column vectors. The k-dimensional projection $\Phi_t^{\mathbf{train}}$ of each $Y_t^{\mathbf{train}}$ is expressed as $P^T Y_t^{\mathbf{train}}$. Through this way, we integrate the spatio-temporal Velib usage patterns within the short-term time frames into a compact k-dimensional projection subspace Ω. During the testing procedure, for the testing day, we firstly project the Velib usage count $Y_j^{\mathbf{test}}$ at the time $j = t-2, t-1$ to the k-dimensional space Ω using the projection matrix P. After that, we calculate the distance in the projection space between the sequence $\left\{\Phi_{j=t-2,t-1}^{test}\right\}$ and the sequences $\left\{\Phi_{j=(l-1)*24+t-2,(l-1)*24+t-1}^{train}\right\}$ at the corresponding time frame $(j = t-2, t-1)$ of each day l in the training data set, in order to identify the p nearest neighbors of the testing day in the subspace Ω. The distance measure between the sequences is defined as summation of cosine distance between the PCA projections:

$$Dis = \sum_{j=t-2,t-1} \frac{\left(\Phi_j^{\mathbf{test}}\right)^T \Phi_{(l-1)*24+j}^{\mathbf{train}}}{\|\Phi_j^{\mathbf{test}}\|_{L^2} \|\Phi_{(l-1)*24+j}^{\mathbf{train}}\|_{L^2}} \tag{7}$$

where $\|\|_{L^2}$ denotes the L_2 norm of vector.The bicycle demand $X_{:,t}^{out}$ of all stations at the time t on the testing day is predicted as the average of the bicycle demands $X_{:,t}^{out}$ at the corresponding time t of the p nearest neighboring days in the training set. The PCA projection conserves global characteristics of Velib usage over the whole network. Nearest neighbor searching in the PCA space considers similarity of spatio-temporal Velib usage patterns between the historical records and the testing sample. The final prediction is a linear combination of the historical bicycle demands with similar precedent spatio-temporal Velib usage pattern. This scheme achieves to predict

Table 1: Prediction accuracies of the three methods

Prediction method	Average prediction error	Variance of prediction error
Station-wise poisson regression	2.05	15.6
NN+PCA	1.47	3.80
The proposed method	1.45	3.65

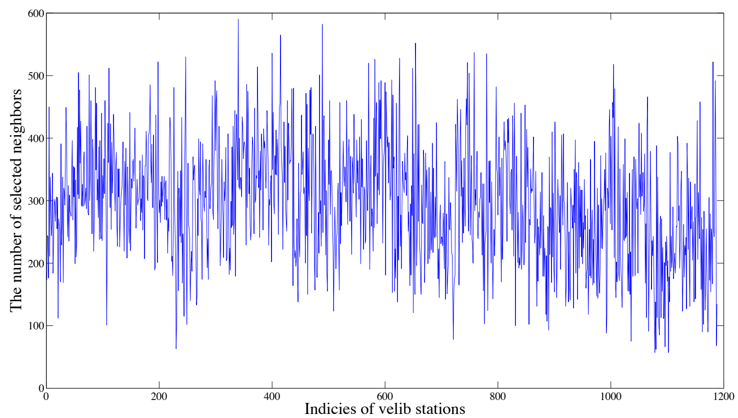
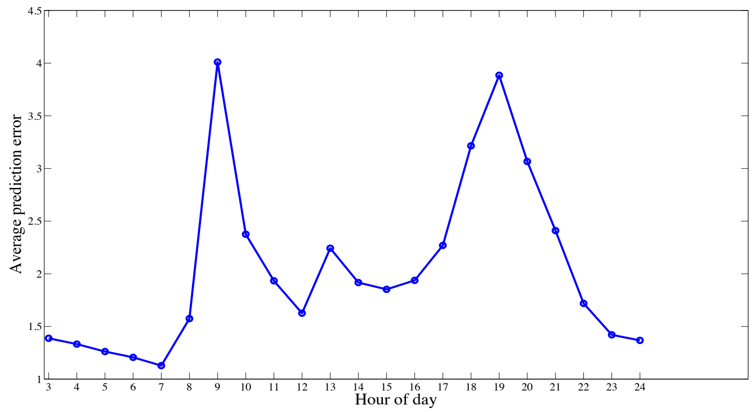


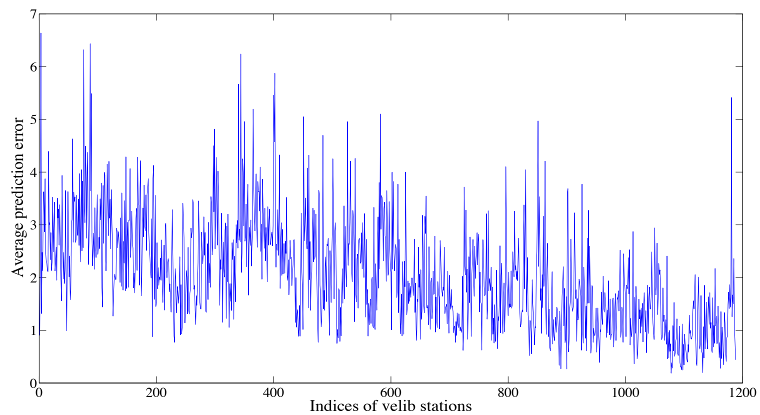
Figure 5: The number of selected neighbors for each station

the Velib demands at all stations simulatensouly. In the followings, we note this scheme as $NN + PCA$ for short. For fair comparison, we adjust the number of the nearest neighbors in prediction to achieve the best performance.

As we can see in Table 1. The proposed negative binomial regression model achieves superior performances to $NN + PCA$ and the station-wise poisson regression model with respect to both mean and variance of prediction error. $NN + PCA$ performs much better compared with the station-wise poisson regression. The experimental results are consistent with the original expectation. Both the proposed spatio-temporal network based prediction method and the $NN + PCA$ make full use of the short-term spatio-temporal Velib usage patterns in constructing the temporal forecast model. The station-wise poisson regression model only depends on station specific Velib usage records. The former two gains more predictive information from the investigated spatio-temporal Velib usage patterns to narrow the variance of Velib demand estimation, making the prediction more close to the underlying values. The network structure learning procedure extracts the prediction-oriented spatio-temporal correlation structure with respect to each station. In contrast, $NN + PCA$ conserves only global spatio-temporal usage patterns. This global information is corase and is not tailored for temporal prediction of each local station. Thus $NN + PCA$ performs less accurately than the proposed method. Figure 5 illustrates the the number of neighbors in X_{Δ_2} for each station. As illustrated, the proposed method benefits from the sparsity-introducing regularization to construct a sparse linkage structure in the spatio-temporal



(a) Average prediction error per hour



(b) Average prediction error per station

Figure 6: Hour-wise and station-wise average prediction error

network. 295 neighbors are selected for each station on average. This sparse linkage structure is helpful in selecting the really useful spatio-temporal correlation of the Veilb usage and improving the computational efficiency of prediction.

Figure 6(a) and Figure 6(b) illustrate the variation of average prediction error for each hour and each station respectively by performing the proposed method in the training/testing process. As shown in Figure 6(a), about ten percent of the all stations has distinctively larger prediction errors than the others. They correspond to the stations around transportation hubs, such as railway stations and places of interests in Paris. Velib usage at those stations are easily affected by the social-economic factors, such as type of the day (week-end, public holiday or common working days) and special events (accidents or adjustment of public transportation modes). Figure 6(b) shows the variation of prediction error corresponding to different hours of day. We can find that there are two peaks of prediction errors in 24 hours of one day. One is centered around

10 am, the other is around 19 pm. These two peaks are consistent with the peaking hour of public transportation. Large travel demand in Paris during the peaking hour increase the variance of Velib usage counts globally in Velib system.

6 Conclusion

In this paper, we aim to predict short-term Velib service demand variations at each station of Velib system based on historical Velib usage records. The simultaneously prediction for all stations is formulated as a spatio-temporal network filtering process. Given the historical Velib usage records as the input set of the network, the linkage structure of the network represents the spatio-temporal correlation structure between the historical information and the prediction target that is highly relevant with the prediction goal. A properly configured network linkage structure will give the accurate prediction efficiently. Therefore, the learning of the underlying network structure plays the key role in this work.

To achieve this goal, we propose to integrate a count data regression model and a sparsity-introducing regularization, named $L1$ regularized negative binomial regression, to identify the most relevant spatio-temporal Velib usage patterns with the prediction target at each station. This procedure thus provides a sparse estimation of the network linkage structure. Due to the count data regression component, the identified linkage is designed to suit the goal of accurate temporal prediction. Benefitted from the $L1$ based sparsity regularization, the derived linkage structure is compact, in order to remove the irrelevant and redundant information from the constructed prediction model. Experiments on massive amounts of Velib usage records in the large-scale urban area verify the superior forecast power of the proposed method. Besides, we also show that the identified spaio-temporal correlation of Velib usage records is consistent with daily Velib usage behaviors in the Velib system. This confirms the capability of the proposed method in describing the intrinsic rules of short-distance Velib travels in the city.

Acknowledgement

The authors wish to thank Franois Prochasson (Ville de Paris) and Thomas Valeau (Cyclocity-JCDecaux) for providing Velib data.

References

[1] APUR. Etude de localisation des stations de vlos en libre service. rapport. Technical Report 349, Atelier Parisien d'Urbanisme, December 2006.

[2] P. Borgnat, C. Robardet, J.-B. Rouquier, Abry Parice, E. Fleury, and P. Flandrin. Shared Bicycles in a City: A Signal processing and Data Analysis Perspective. *Advances in Complex Systems*, 14(3):1–24, June 2011.

[3] S. Boyd and Vandenberghe. L. *Convex Optimization.* Cambridge University Press, 2004.

[4] H. Bttner, J. Mlasowky, T. Birkholz, D. Groper, a.C. Fernandez, Emberger G., and M. Banfi. Optmising bike sharing in european cities, a handbook. Technical report, Intelligent Energy Europe Program (IEE, OBIS projext), August 2011.

[5] Yang J. C. Yan S.C. Fu Y. Cheng, B. and S. T. Huang. Learning with l1-graph for image analysis. *IEEE Transactions on Image Processing*, 19(4):858–866, 2010.

[6] B. V. Dasarathy. *Nearest neighbor (NN) norms: NN pattern classification techniques*. IEEE Computer Society Press, 1991.

[7] P. De Maio. Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation*, 12(4):41–56, 2009.

[8] L. Dell'Olio, A. Ibeas, and J. L. Moura. Implementing bike-sharing systems. In *ICE - Municipal Engineer*, volume 164, pages 89–101. ICE publishing, 2011.

[9] Hastie T. Friedman, J. and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9:432–441, 2008.

[10] J. Froehlich, J. Neumann, and N. Oliver. Sensing and predicting the pulse of the city through shared bicycling. In *21st International Joint Conference on Artificial Intelligence, IJCAI'09*, pages 1420–1426. AAAI Press, 2009.

[11] Neumann J. Froehlich, J. and Nuria. Oliver. Sensing and predicting the pulse of the city through shared bicycling. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 1420–1426. Morgan Kaufmann Publishers, San Francisco, USA, 2009.

[12] J. M. Hilbe. *Negative Binomial Regression: 2nd Edition*. Cambridge, 2011.

[13] Neal Lathia, A. Saniul, and L. Capra. Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies*, 22:88–102, June 2012.

[14] J.R. Lin and T. Yang. Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E: Logistics and Transportation Review*, 47(2):284–294, 2011.

[15] M. Meinshausen and P Buhlmann. High dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34(3), 2006.

[16] P. Vogel and D.C. Mattfeld. Strategic and operational planning of bike-sharing systems by data mining - a case study. In *ICCL*, pages 127–141. Springer Berlin Heidelberg, 2011.

[17] Ravikumar P. Wainwright, J. M. and J. D. Lafferty. High-dimensional graphical model selection using l1-regularized logistic regression. In *20th Neural Information Processing Systems, NIPS 2006*, Vancouver, Canada, 2006.

[18] Rutimann P. Xu M. Zhou, S.H. and P Buhlmann. High-dimensional covariance estimation based on gaussian graphical models. *Journal of Machine Learning Research*, 12:2975–3026, 2011.