2012 IEEE 13th International Conference on Mobile Data Management

Cityride: a predictive bike sharing journey advisor

Ji Won Yoon
IBM Research
Dublin, Ireland
Email: jiwon.yoon@ie.ibm.com

Fabio Pinelli
IBM Research
Dublin, Ireland
Email: fabiopin@ie.ibm.com

Francesco Calabrese
IBM Research
Dublin, Ireland
Email: fcalabre@ie.ibm.com

***Abstract*—In this paper, we present a personal journey advisor application for helping people to navigate the city using the available bike-sharing system. For a given origin and destination, the application suggests the best pair of stations to be used to take and return a city-bike, in order to minimize the overall walking and biking travel time as well as maximizing the probability to find available bikes at the first station and returning slots at the second one. To solve the journey advisor optimization problem, we modeled real mobile bikers' behavior in terms of travel time, and used the predicted availability at every bike station to choose the pair of stations which maximizes a measure of optimality . To develop the application, we built a spatio-temporal prediction system able to estimate the number of available bikes for each station in short and long term, outperforming already developed solutions. The prediction system is based on an underlying spatial interaction network among the bike stations, and takes into account the temporal patterns included in the data. The Cityride application was tested with real data from the Dublin bike-sharing system.**

***Keywords*-journey planner, bike-sharing, mobility analysis, mobile phone application**

I. Introduction

In the context of urban mobility, bike sharing programs have seen a great development during recent years, starting from Europe and quickly expanding to the rest of the world [1]. Bike sharing schemes have been reported to increase the cycling population, and push cities to improve cycle facilities [2]. Moreover, locating bike sharing stations in strategic locations close to public transport services, has allowed improved connectivity to other modes of transit, and so increased overall public transport usage.

Intelligent management of bike sharing systems have seen a great interest within the research community. Researchers have been focusing on the problem of effective planning of bike sharing schemes [3], [4]. Moreover, intelligent techniques to redistribute bikes have been proposed [5].

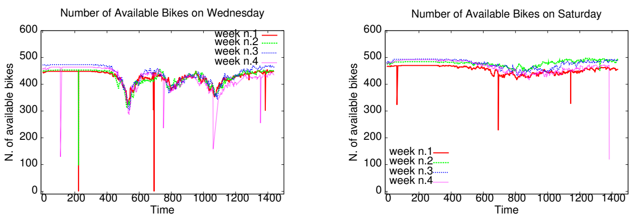
Data on the real time availability of bikes at stations is usually publicly provided by the service operator. There has been an increased interest in applying mobile data management techniques to extract new insights from the bike usage, see for instance [6], [7]. In the case of Lyon, where the service operator provided researchers with fine grained data containing origin and destination of bike trips, further insights could be gained on the flows between stations [8], [9].

In this paper, we focus on bike sharing systems as one of the means of transport within a multimodal transportation network of the city, and deal with the development of a personal journey advisor which can advise citizens on how to navigate the city using the available multimodal network, considering the uncertainties related to the availability of the different services [10]. The idea of a journey advisor is not only to provide the citizen with a sequence of means of transport that could be taken to go from a given origin A to a given destination B in the city, but also to advise on the most effective option in terms of traveling time, waiting time, probability to miss connections, etc. Within this framework, and with reference to the bike sharing system, a static journey planner is quite easy to develop, as the location of bike stations is fixed. However, problems with a static journey planner might arise if suggested stations have no bikes or returning spots. This problem indeed has been found to provide frustration in users if no good redistribution schemes are provided [5]. To give an example, as it will be shown in the next section, bike stations in Dublin can experience long periods of time (3-4 hours per day) in which no bikes or returning spots are available. In particular, we focus on developing the journey advisor application with a prediction system that helps reduce this inconvenience by providing estimates of the availability of bikes at each station, and use them to suggest the fastest and most reliable trip plan. We also leverage publicly available bike-sharing flow data for London and Washington [11], [12] to evaluate travel time and trip length distribution, useful to understand bikers' behavior in terms likelihood to travel a given distance. This lets us model real users' behavior whilst trying to avoid to suggesting too short or too long bike trips.

Existing applications related to bike sharing systems can provide two possible services. On one hand they statically suggest origin/destination station without taking into account real time information; on the other hand there are applications or web sites that in real-time provide the number of available bikes at the stations. However, to the best of our knowledge, they do not provide predicted availability, nor do they allow planning of a journey.

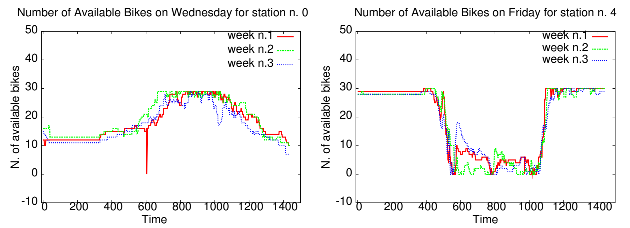
Our system completes the services offered by the existing

978-0-7695-4713-8/12 $26.00 © 2012 IEEE
DOI 10.1109/MDM.2012.16



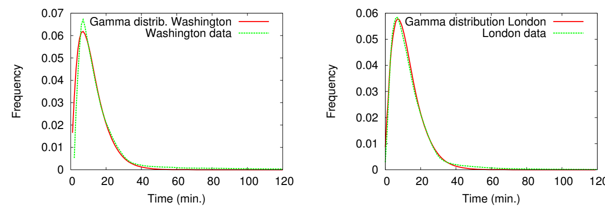
(a) Week day (b) Week-end day

Figure 1. Comparison week and weekend days in terms of total number of available bikes



(a) Station n.1 (b) Station n.5

Figure 2. Different temporal patterns in terms of number of available bikes detected in different bike stations



(a) Washington (b) London

Figure 3. Travel time distribution for the Washington dataset (left) and the London dataset (right)

applications. Indeed, it provides a journey planner which adopts real time information used by a prediction algorithm for the future availability of bikes/returning spots providing a more practical service.

Forecasting the number of rentals has been previously proposed [6], [7], [8]. Our approach differs and outperforms previous ones as: i) we adopt similarity between time series of different stations to improve prediction performance; ii) we leverage the periodicity in the time series to build a temporal model of evolution of the time series.

The paper is structured as follows. Section II describes the dataset used in this paper. Section III describes the proposed journey advisor problem with proposed spatio-temporal prediction systems and how to build the neighbouring structure for the spatial prediction. Section IV presents a real case of study. Section V provides discussion and limitations. Section VI presents conclusive remarks and future work.

II. Data description

The dataset is obtained crawling the web site of the *Dublinbikes* system[1]. The dataset under analysis covers 27 days from January 4th 2011 to January 31st 2011. For every minute, the current number of available bikes is recorded for each of the 42 bike stations of the *Dublinbikes* system.

A. Temporal Pattern in the data

Despite the presence of some noise, the data includes regularities and temporal patterns. In Fig. 1(a) and Fig.1(b) we compare the total number of available bikes during a week day and a weekend day. Firstly, we highlight that the trends depicted during working days are very similar among different weeks, and this is confirmed also looking at the weekend days. We notice that the trends described by the week days and the weekend days are considerably different. Only Fig. 1(a) i.e. week data, presents clear peaks during certain hours of the day. These peaks are missing in Fig. 1(b), i.e week end data.

Moreover, each station presents a different temporal pattern, and therefore this means that people use bike stations in different ways. In particular, station 1 (as numbered on the *Dublinbikes* website), represented in Fig. 2(a), shows some peculiarities. Usually during the night, the station 1 has few bikes available, while during the day many bikes are present. A completely different pattern is shown in Fig. 2(b), where the trend of the station 5 is represented. Many bikes are accessible during the night, while during the day there are long periods where the number of available bikes is very low or no bikes are obtainable.

B. Travel time analysis

As mentioned before, the goal of this paper is to develop a journey advisor taking into account bikers' behavior. The dataset of *Dublinbikes* system does not include information regarding each single trip. In order to estimate the typical bike trip travel time we analyze the freely available data of other bike-sharing systems concerning bike trips performed in London [11] and Washington [12] [2]. Both datasets include information regarding each trip with origin and destination stations, and has been used to compute the distribution of travel times (see Fig. 3). We notice that the graph resembles a gamma distribution in both cases:

$$p(j|i) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} \exp\{-\beta t\} \qquad (1)$$

where t is the duration time for bike travel between station i and j. We estimate the parameters α and β of the gamma distribution through Maximum Likelihood (ML) estimation obtaining $\alpha = 2.55$ and $\beta = 0.19$ for London and $\alpha = 2.25$ and $\beta = 0.19$ for Washington. The relative gamma distributions fit quite well the original data (see Fig. 3).

[1]www.dublinbike.ie

[2]Both systems have the same 30 minutes free pricing scheme as Dublin.

III. Journey Advisor

One of the typical mobility queries from citizens can be which is the most convenient way to go from an origin A to a destination B of a user's journey. Then the goal of the journey advisor, for bike sharing system with N stations, is to find two stations i and j for $i, j \in \{1, 2, \cdots, N\}$ so that a person can use the bike between i and j and he or she walks between A and i and between j and B. The selection of i and j is done as follows:

$$(i^*, j^*) = \arg_{i,j} \max p(i, j|A, B, \mathbf{y}_{1:t}) \tag{2}$$

and we define the target posterior by $p(i, j|A, B, \mathbf{y}_{1:t}) \propto p(B|j)p(i|A)p(j|i)\pi_{av}(i)\pi_{free}(j)$. Here, there are two types of distributions in terms of minimizing distances and available number of bikes. $\pi_{av}(i)$ and $\pi_{free}(j)$ are cumulative density functions of the probability that the ith station has available bikes to take and the jth station has free spots to return respectively. Minimizing distances is a metric for $p(B|j)$, $p(i|A)$ and $p(j|i)$ with different modes. We model the distributions of the travel time for 'walk' and for 'bike' by exponential distribution and gamma distribution (as discussed in section II) respectively:

$$\begin{aligned} p(B|j) &= \lambda \exp\left(-\lambda \Delta^{walk}_{(j,B)}\right), \\ p(i|A) &= \lambda \exp\left(-\lambda \Delta^{walk}_{(A,i)}\right) \text{ and} \\ p(j|i) &= \frac{\beta^\alpha}{\Gamma(\alpha)} {\Delta^{bike}_{(i,j)}}^{\alpha-1} \exp\left\{-\beta \Delta^{bike}_{(i,j)}\right\} \end{aligned} \tag{3}$$

where $\Delta^c_{(a,b)}$ represent the duration of time between site a and site b via mode type $c \in \{walk, bike\}$. The other type is cumulative density functions for taking and returning the bikes:

$$\begin{aligned} \pi_{av}(j) &= \left[1 - p\left(y^{(i)}_{t+\Delta^{walk}_{(A,i)}} = 0|\mathbf{y}_{1:t}\right)\right] \\ \pi_{free}(j) &= \left[1 - p\left(y^{(j)}_{t+\Delta^{walk}_{(A,i)}+\Delta^{bike}_{(i,j)}} = y^{(j)}_{\max}|\mathbf{y}_{1:t}\right)\right] \end{aligned} \tag{4}$$

where $y^{(j)}_{\max}$ is the total number of bikes at the jth bike station. The probabilities in the formulas can be obtained using one of the prediction algorithms presented in the next section.

A. Prediction

In the journal advisor system, one of the most important parts is developing prediction algorithms to infer the precise number of available bikes in each station n-minutes ahead rather than the binary status of the bike availability. The awareness of the precise number of available bikes can be useful for both operators and final users. The former can use this information to better evaluate the system performance, the latter can use it to have a more accurate knowledge of the current status of a station.

There are several approaches to achieve this goal. Recently, [6] showed an AutoRegressive Moving Average (ARMA) model to forecast the available number of bikes and [7] showed a simple Bayesian network based prediction algorithm. However, both approaches present some limitations. First of all, ARMA models are only useful for stationary signals so we need to improve the model as the mean and variance of the bike availability signals are varying over time, bringing non-stationarity. Although the naive Bayesian prediction (classification) based on Bayesian network [7] seems very simple but useful in bike prediction, the authors used small number of classes according to percentages (e.g. 5 classes: $0\% \sim 20\%, \cdots, 80\% \sim 100\%$) rather than predicting the actual available number of bikes, which instead would be needed for accurate journey planning if groups of more than one user are interesting in planning a trip. The other issue in Bayesian Network forecasting in [7] is that the number of conditional factors is too small so that the prediction is not much better than simple Last Value (LV) prediction[3]. In general the performance of a Bayesian Network can be improved by adding more factors, but this leads to an exponential size of the probability table and thus to time consuming simulations. In addition, Both approaches (ARMA and Bayesian Network) have considered each station as independent. We, instead, try to leverage the clear correlation between stations (bike decrease in an origin station are correlated to increase in a destination station, close-by stations usually have similar usage) to improve prediction performances.

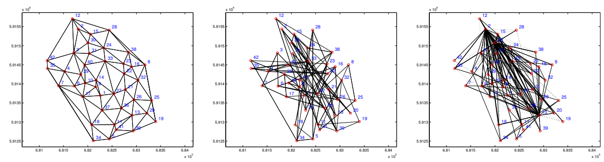
Therefore, we develop several algorithms to address the above limitations. We propose a mathematical model for handling non-stationarity of the data by using AutoRegressive Integrated Moving Average (ARIMA) model. One of our contributions is to deal with spatio-temporal prediction by using signals from neighboring stations and seasonal trend. In order to inlcude this information, we need to modify the ARIMA model by

$$\begin{aligned} \Delta_r y^{(a)}_t &= \sum_{i=1}^{p} \alpha_i \Delta_r y^{(a)}_{t-i} + \alpha_0 + \sum_{n \in ne(a)} \sum_{j=1}^{p} \gamma_{n,j} \Delta_r y^{(n)}_{t-j} \\ &\quad + \sum_{s=1}^{N_s} \rho_s \Delta_r y^{(a)}_{t-sg} + \sum_{j=1}^{q} \beta_j \epsilon^{(a)}_{t-j} + \epsilon^{(a)}_t \end{aligned} \tag{5}$$

where $\Delta_r \mathbf{y}_t$ denotes the rth differencing of the $\mathbf{y}_t$ to impose stationarity and $\epsilon_t \sim \mathcal{N}(\epsilon_t; 0, \sigma^2)$. In this equation, α_i, β_j, $\gamma_{n,j}$ and ρ_s are the coefficients of AR, MA, spatial pattern and seasonal trend respectively. We estimate the parameters via iterative ML (or REML[4]). Here, g is the seasonal period and we simply set this to one week and $ne(a)$ is the set of the neighbour stations of the ath station given a particular neighbouring structure which will be described below. For simplicity, we simply used $N_s = 1$ in this paper.

[3]LV is a simple algorithm for which the n-minutes ahead prediction is only replaced by the last observed value.

[4]REML is statistical term for Restricted Maximum Likelihood, which is unbiased estimator.



(a) Voronoi (b) KNN_{sim} $(K = 3)$ (c) Linear regression

Figure 4. Reconstruction of spatial interaction of Dublinbikes: the locations (latitude and longitude) of 42 bike stations and their neighbouring network. In Linear Regression model, the solid and dot lines represent significantly positive and negative connection respectively.

B. Exploring the neighboring structure

As we can see in spatio-temporal ARIMA, we are using spatial information obtained from the neighboring stations. Since there is no explicit interaction network between bike stations, we explore different network topologies to determine which one is most effective from a prediction point of view. There are several potential approaches to find underlying neighbors. In the following, we describe three methods.

1) Voronoi tessellation based: We build a Voronoi tessellation to reconstruct the neighboring structure, which assigns the neighbors automatically by using the geometric information as shown in figure 4(a). Voronoi tessellation is generated via location closeness.

2) K-nearest Similar pattern based (KNN_{sim}): The signals (the number of available bikes) from the bike rental system may be well correlated not by the signals of nearest stations but by similar patterned stations. Therefore, we compared the sequences of the number of available bikes of all stations and then we applied KNN approach to find neighbors as shown in figure 4(b).

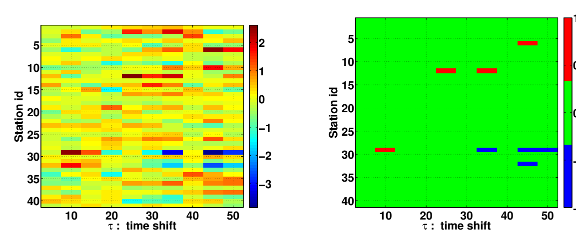
3) Linear Regression based: Such a similar pattern based KNN approach can detect similar pattern or trend for the paired sequences. However, in a bike sharing system there is another interesting point to be considered. If a person borrows a bike from station i and then he can return the bike at station j. That is, there are two different correlations, positive (similar pattern) and negative (opposite pattern). In order to detect both patterns we calculate the correlation coefficients via Generalized Linear Regression model.

$$\mathbf{S}_\tau^{(a)} = \mathbf{R}_\tau \mathbf{c}_\tau + \mathbf{v}_\tau^{(a)} \tag{6}$$

where $\mathbf{S}_\tau^{(a)} = \mathbf{y}_{\tau+1:t}^{(a)'}$, $\mathbf{c}_\tau = \rho_{1:N}$, $\mathbf{R}_\tau = \left[\mathbf{1}_{1:t-\tau}', \mathbf{y}_{1:t-\tau}^{(1)'}, \mathbf{y}_{1:t-\tau}^{(2)'}, \cdots, \mathbf{y}_{1:t-\tau}^{(N)'}\right]$ and $\mathbf{a}_{1:N}^{(\tau)}$ is a $N \times 1$ vector for shift window τ and $\mathbf{v}_\tau \sim \mathcal{N}(\mathbf{v}_\tau; \mathbf{0}, \Omega)$ and $\Omega = \sigma_c^2 \mathbf{I}_{(t-\tau)\times(t-\tau)}$. Given this assumption, we can see that to find the optimal weights $\mathbf{c}_\tau$ and the unknown variance are iteratively calculated via a Generalized Least Square scheme. Thus, we have $\mathbf{E}(\mathbf{c}_\tau) = (\mathbf{R}_\tau' \Omega^{-1} \mathbf{R}_\tau)^{-1} \mathbf{R}_\tau' \Omega^{-1} \mathbf{S}_\tau$, $\mathbf{V}(\mathbf{c}_\tau) = (\mathbf{R}_\tau' \Omega^{-1} \mathbf{R}_\tau)^{-1}$ and $\sigma_c^2 = \frac{1}{t-\tau}(\mathbf{S}_\tau - \mathbf{R}_\tau \mathbf{c}_\tau)'(\mathbf{S}_\tau - \mathbf{R}_\tau \mathbf{c}_\tau)$. Now, we can make a map for weights with varying time shift τ. For instance, we have a map for the ith site by $\mathbf{C}_a = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_\tau]$. Therefore, we have 3 dimensional maps for the all sites by $\mathbf{C} = \{\mathbf{C}_a\}_{a=1}^N$ and $\mathbf{C}$ is $N \times (t-\tau) \times (N-1)$ size. The correlation coefficient map for the 7th station in Dublin is shown in figure 5. Once we obtain $\mathbf{C}_a$, then we need to process it to detect significant neighbors. We simply use a clustering algorithm based on histogram. Here, two assumptions are employed: (1) most stations are not connected so the coefficients will be close to zero and (2) stations with significant influence will have a large negative (highly correlated negative factor) or large positive value (highly correlated positive factor). With these two assumptions, we first calculate the mean (μ_c) and standard deviation (σ_c) of $\mathbf{C}_a$ for $a \in \{1, 2, \cdots, N\}$. Afterwards, we partition it into three clusters: significant positive , not significant and significant negative as follows:

- **Significantly positive:** $\Psi_+ = \{r | r > \mu_c + 3\sigma_c\}$,
- **Significantly negative:** $\Psi_- = \{r | r < \mu_c - 3\sigma_c\}$ and
- **Not significant:** $\Psi_{\cdot} = \{r | \mu_c + 3\sigma_c \le r \le \mu_c + 3\sigma_c\}$ for all $r \in \mathbf{C}_a$.

As an example, the classified map for station no. 7 is shown in figure 5-(b) clustering the actual coefficient map in Fig. 5-(a) by the previously defined rules. Red and blue slots represent the significantly positive and negative stations to a particular station with varying time shift.



(a) Regression Map (b) Classified Map

Figure 5. The correlation map of 7th station and its classified map via Linear Regression approach: x-axis and y-axis represent the time shift (τ) and the 42 stations respectively. In the classified map (b), red and blue slots show the significantly positive and negative stations.

Given the classified maps for all stations, we can obtain the neighbors as shown in figure 4(c).

IV. Application to Real-time Dublinbikes data

We tested and compared the performance of our prediction algorithms for one month with other known approaches: Last Value (LV), Historical Mean (HM), Median filter and ARMA [6]. We have used 10 days for the training set and for learning the model parameters of variants of ARIMA model, with a sampling time of 5 minutes.

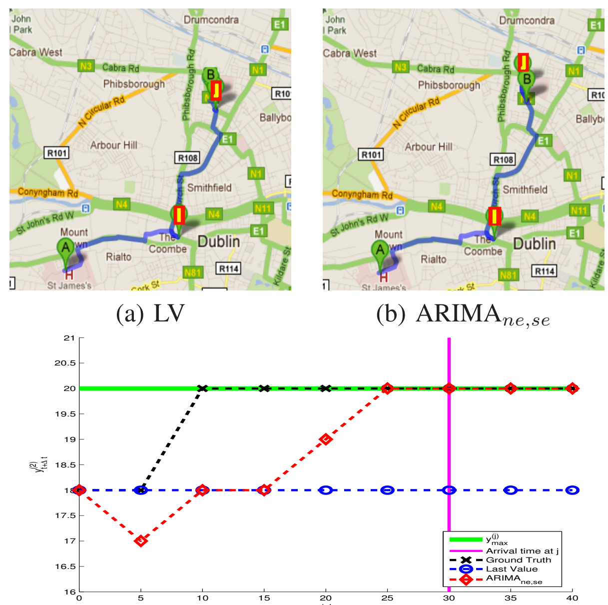
In Table I we report the performance comparison of different algorithms considering various neighboring structures for both short (5 minutes ahead) and long term (60 minutes ahead) predictions. We used the Root Mean Square Error (RMSE) as metric for performance assessment. We notice that the ARIMA-based models outperform the previously proposed methods. Moreover, a peculiar trend emerges if we consider the different neighboring structures included in the model. For instance, the ARIMA algorithms with complete neighbors structure have high RMSE compared to others. The training dataset used is not large enough to properly estimate myriads of unknown parameters, where the number of unknown parameters increases according to the complexity of the neighbouring structure. In addition, the approaches using Voronoi and Regression methods show better performances compared to KNN$_{sim}$ and the complete neighbors. One of the most interesting results is that while ARIMA$_{ne}$ outperforms ARIMA$_{ne,se}$ in all neighbor structure for the short term prediction, ARIMA$_{ne,se}$ becomes more accurate than ARIMA$_{ne}$ in the long term prediction. This is because the neighbors and past observations are important factors to provide accurate short term prediction. However we can obtain more informative and significant information from the temporal trends than the neighbours and previous values for the long term prediction.

	$5min$	$60min$	neighbour
LV	2.04 ± 0.63	4.23 ± 1.27	
HM	1.20 ± 0.34	3.57 ± 1.15	
Median	1.27 ± 0.36	3.63 ± 1.18	
ARMA(3,3)	0.92 ± 0.24	3.50 ± 1.09	
ARIMA$_{se}$(3,1,3)	0.92 ± 0.23	3.48 ± 1.09	
ARIMA$_{ne}$ (3,1,3)	1.15 ± 0.67	3.80 ± 1.23	Complete
	0.91 ± 0.24	3.49 ± 1.12	Voronoi
	0.91 ± 0.24	3.50 ± 1.12	KNN$_{sim}$
	$\mathbf{0.91 \pm 0.23}$	3.50 ± 1.11	Regression
ARIMA$_{ne,se}$ (3,1,3)	1.23 ± 0.77	3.91 ± 1.39	Complete
	0.92 ± 0.24	$\mathbf{3.47 \pm 1.09}$	Voronoi
	0.93 ± 0.24	3.48 ± 1.11	KNN$_{sim}$
	0.92 ± 0.24	$\mathbf{3.47 \pm 1.09}$	Regression

Table I
COMPARISON OF ROOT MEAN SQUARE ERROR (RMSE) OF THE SEVERAL PREDICTION ALGORITHMS WITH VARYING NEIGHBOURING STRUCTURE

To show how prediction inaccuracy influences the performance of the journey advisor, we show an example of a query by A ='St. James hospital, Ireland (53.3394, -6.2958)' and B ='Berkeley Street, Dublin, Ireland (53.35757, -6.2690)' and we compare the suggested plans obtained using two different prediction algorithms: the simple LV (which corresponds to the real time bike availability data currently available to Dubliners using the official website) and our proposed ARIMA$_{se,ne}$. In this example, we set α and β as estimated from the London bike dataset as mentioned in data section, and $\lambda = 1^5$ Considering a particular time of the day (21:41pm, 17th, Jan., 2011), LV outputs the shortest path via the 7th station for i and the 2nd station for j, instead our proposed ARIMA$_{se,ne}$ recommends a different returning station $j = 12$, even if it results in a longer path (see Figure 6). The reason for the different paths is due to the fact that the LV method is not able to predict that there will not be free returning spots at station 2 by the time the user will arrive, estimated to be 30 minutes. Indeed, if we compare predicted availability at station 2 after 30 minutes (see figure 6-(c)) we see that LV wrongly predicts 2 available returning spots. Conversely, our ARIMA$_{se,ne}$ method is able to predict that station 2 will be full, and so the advisor will select another close-by station with free returning spots.

[5]As no quantitative data was available to estimate λ, it has been chosen based on the authors experience. Further discussion on how this value can be fine-tuned are given in section V.



(a) LV (b) ARIMA$_{ne,se}$

(c) the predicted number of bikes at the 2nd station

Figure 6. The best paths recommended by the journey advisor, using LV and ARIMA$_{ne,se}$ methods: from 'St. James Hospital, Dublin' to 'Berkeley Street, Dublin': the best paths by LV (top left) and by ARIMA$_{ne,se}$ (top right) at query time 21:41pm, 17th, Jan., 2011. (c) Available number of bikes at station number 2, starting at 21:41pm ($\Delta t = 0$), and predicted future availability using LV and ARIMA$_{se,ne}$.

V. DISCUSSION

In this section we report several considerations regarding the proposed application.

- The optimization problem tries to minimize expected uncertainty-aware travel time. The optimization function could be further developed to take into account, for instance, the travel cost or the risk to arrive at destination within a specific time. A threshold could be added so that, for very short trips, no bike trip is suggested. Moreover, another threshold could be added for very long trips, when alternative solutions using other transport modes should be considered.

- *Bike redistribution strategy*: given the currently available data, we can only estimate when trucks take or add bikes to stations to redistribute them (by looking at many bikes taken or added in the same minute) and use this estimation as part of the prediction system. As for the bikers' dynamics, our application so far does not take into account that providing predicted availability to service operators could push changes in redistribution strategies. A city wide manager application developed together with the Dublin City Council [13] is being complemented with real time and predicted bike availability, to help the city's department of transportation to manage the multi-modal transportation network. This will allow us to evaluate the impact of the application to the bike-sharing management system.
- The available dataset for Dublin does not have origin-destination bike flow data, as it is based only on publicly available data. This data could improve the system as we could infer when a bike has been taken, and its destination with some probability. Other related datasets, such as mobile phone data [14] could be used to infer overall origin-destination flow and statistics leveraged to infer bike origin-destination flow share.
- The application could be extended to allow for that real time adaptation of the predictions, that can trigger the application to recommend the user (even during his/her bike trip) to change destination station as predicted availability changes. It could happen that the suggested station might switch over time between two close-by stations, as their estimated travel time might be very close. To avoid this, close-by stations could be clustered together so that predictions could be done for a group of stations instead of individual ones. Once the user arrives in the vicinity of the group of stations, the application could decide which station, among the ones in the group, minimizes the optimization cost function, and so suggests the best one.

VI. Conclusion

In this paper we have proposed a personal journey advisor application for navigating the city using the bike-sharing system. To develop the application, we built a spatio-temporal prediction system based on ARIMA, extended to include seasonal trends and regression-based spatial correlations, outperforming already developed solutions. We modeled the uncertainties related to the prediction system and real bikers' behavior in terms of travel time, and used them to solve the journey advisor optimization problem. The application was developed and tested on data from the Dublin bike-sharing system. As future work, we are planning on testing prediction performance on larger datasets and on different cities, and on evaluating prediction systems as function of the features of each station. We are also working on integrating other transportation modes to develop a multi-modal journey advisor.

References

[1] S. A. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in Europe, the Americas, and Asia past, present, and future," *Journal of the Transportation Research Board*, vol. 2143, pp. 159–167, 2010.

[2] P. deMaio, "Bike-sharing: History, impacts, models of provision, and future," *Journal of Public Transportation*, vol. 12, pp. 41–56, 2009.

[3] G. Krykewycz, J. Rocks, B. Bonnette, and F. Jaskiewicz, "Defining a primary market and estimating demand for major bicycle-sharing program in philadelphia," *J. of the Transportation Research Board*, vol. 2143, pp. 117–124, 2011.

[4] J.-R. Lin and T.-H. Yang, "Strategic design of public bicycle sharing systems with service level constraints," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 2, pp. 284–294, 2011.

[5] R. Nair and E. Miller-Hooks, "Fleet Management for Vehicle Sharing Operations," *Transportation Science*, 2010.

[6] 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, 2010.

[7] J. Froehlich, J. Neumann, and N. Oliver, "Sensing and predicting the pulse of the city through shared bicycling," in *International Joint Conference on Artificial Intelligence (IJCAI-09)*, 2009.

[8] P. Borgnat, C. Robardet, J. Rouquier, P. Abry, P. Flandrin, and E. Fleury, "Shared bicycles in a city: A signal processing and data analysis perspective," *Adv. in Complex Systems*, 2011.

[9] P. Jensen, J.-B. Rouquier, N. Ovtracht, and C. Robardet, "Characterizing the speed and paths of shared bicycles in lyon," *Transportation Research Part D: Transport and Environment*, vol. 15, 2010.

[10] A. Baptista, E. Bouillet, F. Calabrese, and O. Verscheure, "Towards building an uncertainty-aware personal journey planner," in *IEEE Intelligent Transportation Systems Conference*, Washington, DC, October 2011.

[11] "Transport for London," Barclays Cycle Hire statistics. [Online]. Available: http://www.tfl.gov.uk/businessandpartners/syndication/

[12] "Virginia Department of Rail and Public Transportation." [Online]. Available: http://capitalbikeshare.com/trip-history-data

[13] L. Gasparini, E. Bouillet, F. Calabrese, O. Verscheure, B. O'Brien, and M. O'Donnell, "System and Analytics for Continuously Assessing Transport Systems from Sparse and Noisy Observations: Case Study in Dublin," in *IEEE Intelligent Transportation Systems Conference*, 2011.

[14] F. Calabrese, G. DiLorenzo, L. Liu, and C. Ratti, "Estimating Origin-Destination Flows Using Mobile Phone Location Data," *IEEE Pervasive Computing*, 2011.