On the crowdsourced repositioning scheme for public bike sharing systems

I-Lin Wang †
Department of Industrial and Information Management
National Cheng Kung University, Tainan, Taiwan
Tel: (+886) 6-2757575-53123, Email: ilinwang@mail.ncku.edu.tw

Chen-Tai Hou
Department of Industrial and Information Management
National Cheng Kung University, Tainan, Taiwan
Tel: (+886) 6-2757575-53123, Email: vjack070707@gmail.com

Abstract. The public bike sharing systems might be the most popular applications of shared economy in transportation nowadays. The convenience of such a system depends on availability of bikes and empty racks. One of the major challenges in operating a bike sharing system is to reposition bikes between rental sites to keep good bike inventory in each station at any time. To this end, most systems hire trucks to load bikes from stations of fewer empty racks to stations of fewer bikes. We have analyzed such a common-practiced repositioning scheme and will show its ineffectiveness. To provide a repositioning strategy that provides better service, we proposed a crowdsourced repositioning scheme. In particular, we first analyze the historical rental data by random forest algorithm to identify important factors towards demand forecasting. Then, by setting time tags for every 30mins as a period, we propose a minimum cost network flow model to calculate optimal voluntary rider flows for each period, based on the current bike inventory adjusted by forecasted demands. Assuming some of the voluntary riders do follow our instructions to ride bikes from specific origins to destinations, the repositioning operations can be processed with much better effectiveness than conventional trucks.

Keywords: bike sharing, crowdsourced, repositioning, minimum cost flow, random forest

1. INTRODUCTION

The recent boom in the sharing economy has led to new business and changed how people live in several ways. For example, the public bike systems in many metropolitan areas have help people to have more convenient access to the public transportation system such as MRTs, trains, or even buses. Although the business model of bike sharing systems are still arguable in making profits, by July 2016, more than 1000 cities worldwide have installed bike sharing systems, and approximately more than 1.3 million public bikes and pedelecs (i.e., pedal electric cycle) are in use. Indeed, a bike sharing system is a perfect means to serve for the first and last mile connection to public transportation systems. By deploying the rental sites in a suitable density (e.g., each site is within 300m-500m range to another site), one can easily rent a bike from a site nearby, ride and return it to another site to take MRT, exit the MRT station and ride another bike to a site closer to his/her destination.

The success to the bike sharing systems depends on the qualities of service in the following aspects: (1) ease to access a rental site, and (2) ease to take or return a bike. The first aspect is a long term strategic level network design decision, where the locations and density of rental sites should be carefully determined to provide easy accessibility to the users and good connection to the public transportation systems. On the other hand, the second aspect contains series of short term tactical or operational level decisions, where appropriate sizes of bike fleet should be deployed in different time. To this end, the system managers have to reposition bikes between sites to meet the expected bike rental or return demands. This contains several challenging tasks: (1) a good prediction to these demands, i.e., the timing and number of bikes to be taken or returned by customers at each rental site; (2) an optimal bike inventory in each site at any time; and (3) how to effectively reposition bikes to meet the optimal inventory levels in (2). Note that the failure to reposition these bikes would result in shortages in bikes or empty racks, and could not meet the demands to take or return bikes whenever necessary, which would damage the service qualities, discourage the use of such systems.

Figure 1: A repositioning truck in YouBike (the bike sharing system in Taipei city)

To the best of our knowledge, most current bike sharing systems hire trucks to reposition bikes. These trucks (see Figure 1) are usually of small or medium size for easy parking and moving in metropolitan area, and can carry around up to 20 bikes. To load a bike to the truck a staff first checks out a bike from a rack, and then move it to the truck. According to our survey, loading or unloading a bike at least takes 30 secs. A truck may take around 15 min to finish at most 30 loading/unloading tasks in each rental site in each stay. Driving to another site may take about 15 min on average. As a result, one truck may conduct 30*2=60 loading/unloading operations per hour. Assuming there are N trucks, and each truck works for 18hr (e.g., 06:00 to 24:00) without rest, we can then estimate the upper bound on the number of loading/unloading operations within one day to be $60*N*18=1080N$.

Taking the 2014 rental data of YouBike for example, there are around 200 rental sites, N=10 trucks, and there are about 5000 bikes in total. Each truck can visit 2*18=36 sites, and at most a site is visited by 360/200=1.8 trucks on average a day. In other words, at any moment, only 10/200=5% sites can be served by trucks.

In addition, 1080*10/2=5400 bikes are repositioned in one day, since each bike repositioning involves one loading and one unloading process, respectively. Note that YouBike in 2014 has around 40000 daily rentals on average, which means the effects of the 5400 repositioned bikes can be canceled out by 5400/40000=13.5% (around 1/8) users.

All of the above effectiveness analysis for the reposition trucks indicates that it is very ineffective to reposition bikes by trucks. The only way to improve the effectiveness is to increase the number of trucks, yet this would cause more air pollution and traffic jams and contradict the very underlying philosophy of bike sharing systems—to reduce the use of fueled vehicles and carbon emission. To improve the repositioning effectiveness without more repositioning trucks, here we propose a novel *crowdsourced repositioning scheme* to reposition bikes with less efforts, air pollution, and costs. Even better, our proposed scheme may in fact strengthen the loyalty of users to the system.

In particular, we first analyze the historical rental data, design visualization tools to help system managers understand the trends of rental. Based on these data, we first solve an integer program (Ideal Inventory Model, IIM) that seeks an optimal bike inventory level for each station in each time period (e.g. every 30 min), assuming we can always reposition bikes whenever necessary. To predict inventory change in near future, we first use the *random forest algorithm* to identify important factors to the rental trends, and then use them to give better rental prediction within next time period. With more accurate prediction in the inventory trends and target inventory levels for each site at each time period, we formulate a linear program called Voluntary Rider Flow Model (VRFM), which is a minimum cost flow problem to be exact, to calculate optimal bike flows for inviting voluntary riders to meet the target bike inventory levels for the incoming time period.

To seek voluntary riders, we suggest to enhance the membership database by adding records of voluntary rides into ordinary historical riding records for each member. To the best of our knowledge, most bike sharing systems even do not store riding historical data for each member. The bike sharing systems should also provide easy access to check out the riding records for each member by websites or applications (APPs) in smart phones. By using our proposed scheme, the system can design some kind of bonus points to encourage members for more frequent rides, and more bonus points for inviting voluntary riders to take repositioning missions. For example, the system can announce or spread out the voluntary missions on the website or APPs, completing each mission can earn some bonus points, which can be cashed out for extra free riding time or gifts. By encouraging volunteer riders to take missions for cashing out bonus, shared vehicles can be simultaneously repositioned at many rental sites. This crowdsourced repositioning strategy would bring at least 3 advantages: (1) prompt responses to reposition bikes in more sites simultaneously, compared with conventional trucks; (2) cost savings in hiring trucks and staffs which in turn reduces the use of fueled vehicles and traffic jams; and (3) increasing the loyalty of members and improving the relationships with other companies that provide services for member to cash out their bonus points. To the best of our knowledge, similar bonus schemes may only have appeared in very few occasions such as encouraging uphill riders in hilly areas, but have not been applied for general-purpose usage. This makes our contribution more significant, since we may be the first to determine voluntary repositioning OD pairs by theoretical mathematical models and algorithms, rather than intuitive marketing techniques.

This paper is organized as follows: Section 1 introduces backgrounds, drawbacks of current repositioning scheme, and advantages of our crowdsourced repositioning scheme;

Section 2 reviews related literature; Section 3 explains the mechanism of IIM for calculating the ideal optimal bike inventory and details and effects of our crowdsourced repositioning scheme VRFM; Section 4 presents the data analysis and computational tests on our proposed models; Section 5 concludes the paper and suggests topics for future research.

2. LITERATURE REVIEW

Here we focus more on the literature of data analysis and repositioning strategies.

Gebhart and Noland (2014) investigated how the weather conditions affect the bike usage trend for the bike sharing system at Washington DC, USA. They found that riders under rain are usually registered users, or those who with private bikes, but not the nonregistered users. Barga et al. (2014) presented an interactive visualization system to display the rental data of Boston, Washington DC, and Chicago in different time, location. Their system can also present the busiest sites at any time. O'brien et al. (2014) list and compare the locations of rental sites, rental data, and weather conditions for 30 bike sharing systems in different time. Sarkar et al. (2015) implemented a similar system but displayed the usage by percentage and grouped rental sites by the usage.

Vogel et al. (2011a,b) calculated the locations for installing bikes and racks by mining the rental data. Montoliu (2012) proposed an algorithm to identify the trend of inventory changes based on the rental data. Froehlich et al. (2009) investigated the demands change in weekday/weekend, the relation between the rental frequencies and locations, and important factors affecting the rentals. They also tested four methods to predict the future demands of next time period based on the current inventory with errors up to 15%. Kaltenbrunner et al. (2010) proposed a prediction model based on Auto-Regressive Moving Average (ARMA). Yoon et al. (2012) also developed an ARMA based model that further considered the seasonal and spatial factors and claimed his model is better than previous ARMA and Bayesian models for the bike sharing system in Dublin.

Rixey et al. (2013) analyzed 3 bike sharing systems in USA, listed factors used for prediction model by multivariate regression analysis. Cagliero et al. (2016) claimed that Bayes Classifiers can do better prediction than the regression-type algorithms for the Citi Bike at New York, and proposed the STation Occupancy Predictor to predict short-term future bike inventories. Recently, Yang et al. (2016) analyzed the rental data of the bike sharing system at Hangzhou, China, and proposed a prediction model for bike rentals by the Random Forest algorithm (RF), which they claimed to have better prediction performance than several other algorithms. They use the bike rental model to estimate the bike returns, similar to the simulation models by Wang and Wu (2016). Here in this paper we will also employ RF to predict incoming demands.

The *static repositioning* problem (Chemla, Meunier and Calvo, 2009; Raviv, Tzur and Forma, 2013; and Benchimol, et al., 2011) investigate how to move bikes at night when there are very few or no demands to meet the target initial bike inventory for each site. On the other hand, the dynamic repositioning problem (Chang, 2010; Hung, 2011; Contardo et al., 2012; and Vogel et al., 2014) calculates the routes for repositioning trucks and number of bikes to be loaded or unloaded in each site. The integer programming models for solving these repositioning problems usually could not deal with cases of more than 60 rental sites, due to the complexity issues. In addition, several heuristics (Hung, 2011; Contardo et al., 2012; and Vogel et al., 2014) have also been proposed but still have bad performance. Kaspi et al. (2014) gave a reservation scheme that allows users to reserve a bike/car and an empty rack/spot for public bike and car sharing systems. By using simulation, they concluded introducing the reservation scheme could reduce the waiting time for renting/returning a shared vehicle, as expected.

Liao (2012) gave the first crowdsourced repositioning model for dynamic repositioning. Assuming the OD demands for each site and each time period (e.g., 30 min) have been estimated, Liao (2012) added possible voluntary riding OD arcs for each site in each period, and solved a mixed integer program to identify optimal voluntary riding assignments for each site in each period. This model could not deal with real-time voluntary repositioning, since it only uses historical average demands as inputs. In this paper, we have resolved this difficulty. Based on the work of Liao (2012), we propose an ideal inventory model (IIM) that calculates the ideal inventory for each site in the beginning of each time period, and a real-time Voluntary Rider Flow Model (VRFM) that seeks the optimal voluntary riders in each time period.

3. MATHEMATICAL MODELS

3.1 The Ideal Inventory Model (IIM)

The optimal bike inventory for each rental site may vary at any time, depending on the dynamic rental demands. To simplify the problem, we use 30 min as the length of a time period (e.g., 06:00 - 06:30 - 01:00 - 01:30 -…- 23:30 - 24:00), and assume the optimal bike inventory remain the same for any time within the same time period. Assume there are N stations, B bikes, T time periods, U_i empty bikes for site i. Let A denote the set of possible OD pairs. For each site i in period t, let b_i^t and r_i^t represent number of bikes to be taken and returned, respectively. Assume there are at most R^t voluntary riders available in period t, and each voluntary rider who departs at period t would spend Δ_{ij} time periods to reposition a bike from site i to j.

We would like to determine the following variables: in each time period t, x_{ij}^t represents the optimal voluntary rider flows for each OD pair $(i,j)\in A$; for each site i, I_i^t denotes its optimal bike inventory level, ΔU_i^t is the optimal number of bikes exceeding the capacity (total number of racks), and ΔL_i^t is the optimal number of bike shortages. Let ε represent a very small number, the ideal bike inventory model (IIM) can be formulated as follows:

$$\min \sum_{t=1}^{T}\sum_{i=1}^{N}\left(\Delta U_i^t+\Delta L_i^t\right)+\varepsilon\sum_{t=1}^{T}\sum_{(i,j)\in A}x_{ij}^t \qquad \text{(IIM)}$$

$$I_i^t = I_i^{t-1}-b_i^t+r_i^{t-1}-\sum_{(i,j)\in A}x_{ij}^t+\sum_{(j,i)\in A}x_{ji}^{t-\Delta_{ji}}-\Delta U_i^t+\Delta L_i^t$$
$$\forall t=1,\dots,T; i=1,\dots,N \qquad (1)$$

$$\sum_{i=1}^{N} I_i^0 = B \qquad (2)$$

$$\sum_{(i,j)\in A} x_{ij}^t = R^t \quad \forall t=1,\dots,T \qquad (3)$$

$$0\le I_i^t\le U_i,\ \Delta U_i^t\ge 0,\ \Delta L_i^t\ge 0\ \forall t=1,\dots,T; i=1,\dots,N \quad (4)$$

$$x_{ij}^t=0 \quad \forall (i,j)\in A,\ t\le\Delta_{ij} \qquad (5)$$

$$x_{ij}^t\ge 0 \quad \forall t=1,\dots,T;\ (i,j)\in A \qquad (6)$$

The objective function minimizes the mismatched demands (i.e., the surplus or shortage in bikes) as well as unnecessary voluntary riders. Constraints (1) define the flow balance relation for bike inventory in the end of each period and each site. Constraints (2) conserve the total number of bikes, whereas constraints (3) conserve the total number of voluntary riders in each period. Constraints (4)(5)(6) define the ranges of variables.

If there are unlimited voluntary riders available at any time anywhere, then we should always meet the target optimal inventory (i.e., I_i^t) anytime anywhere. We call this target inventory to be the "ideal" inventory since such an inventory would serve the most rental demands, regardless the repositioning costs. Thus we will use this ideal inventory as a target inventory value to achieve for each site and period in the real-time crowdsourced repositioning model VRFM.

3.2 The Voluntary Rider Flow Model (VRFM)

Since IIM assumes the rental data are all known and fixed, we propose the voluntary rider flow model (VRFM), which can be viewed as a partial version of IIM decomposed by periods, to deal with the dynamic real-time rental demand.

In particular, in the beginning of period t, for each site i, let $\bar{b}_i^t$ and $\bar{r}_i^t$ represent the predicted (e.g., by random forest algorithm) number of bikes to be taken and returned, respectively; $\bar{I}_i^{t-1}$ denote the real-time current bike inventory, and I_i^t is the target optimal bike inventory calculated from IIM; $\bar{x}_{ji}^{t^-}$ is the expected number of actual voluntary riders who are currently on the way from site j and expected to arrive at site i in this period. Other parameters are the same as IIM.

We would like to determine $\tilde{x}_{ij}^t$, the optimal voluntary rider flows for each OD pair $(i,j)\in A$, and $\tilde{I}_i^t$, the planned ending inventory in current period t, so that $\tilde{I}_i^t$ is as close to I_i^t as possible with minimum efforts in repositioning. In the beginning of period t, we can form the following linear program VRFM^t:

$$\min \sum_{i=1}^{N}\left|\tilde{I}_i^t-I_i^t\right|+\varepsilon\sum_{(i,j)\in A}\tilde{x}_{ij}^t \qquad (\text{VRFM}^t)$$

$$\tilde{I}_i^t=\bar{I}_i^{t-1}-\bar{b}_i^t+\bar{r}_i^t-\sum_{(i,j)\in A}\tilde{x}_{ij}^t+\sum_{(j,i)\in A}\bar{x}_{ji}^{t^-} \quad \forall i=1,\dots,N \ (7)$$

$$\sum_{(i,j)\in A}\tilde{x}_{ij}^t = R^t \qquad (8)$$

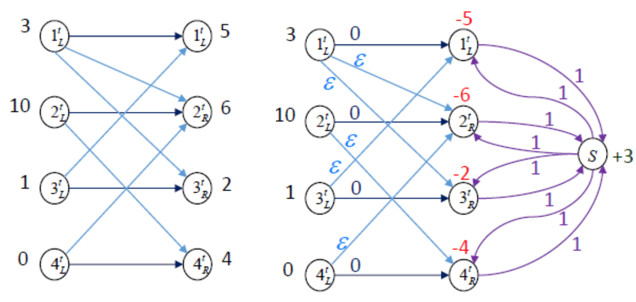
$$0\le \tilde{I}_i^t\le U_i \quad \forall i=1,\dots,N \qquad (9)$$

$$\tilde{x}_{ij}^t\ge 0 \quad \forall\ (i,j)\in A \qquad (10)$$

The linear program VRFM^t in fact corresponds to a minimum cost flow problem as follows: Draw N nodes in the left, with index $i_L^t, i=1,\dots,N$, representing each site at current time t; draw another N nodes in the right, with index $i_R^t, i=1,\dots,N$, representing each site in the next period. For each left node $i_L^t, i=1,\dots,N$, we associate it with value $\bar{I}_i^{t-1}-\bar{b}_i^t+\bar{r}_i^t+\sum_{(j,i)\in A}\bar{x}_{ji}^{t^-}$ to mean its expected inventory before repositioning. For each right node $i_R^t, i=1,\dots,N$, we associate it with value I_i^t to mean its target optimal inventory after repositioning. Then, for each possible OD pair $(i,j)\in A$, we construct an arc (i_L^t, j_R^t) as an *repositioning arc*. We also construct N *inventory arcs* $(i_L^t, i_R^t),\ i=1,\dots,N$.

Figure 2 illustrates a small VRFM example where there are 4 rental sites with expected bike inventory vector as (3, 10, 1, 0), and the target optimal inventory vector as (5, 6, 2, 4), respectively. Note that we know there will be shortage of 3 bikes, since (3+10+1+0)-(5+6+2+4)=-3. Suppose we can only find voluntary riders to move from site 1 to 2, 1 to 3, 2 to 4, 3 to 1, and 4 to 1, respectively. Figure 2(a) illustrates the original graph, which then can be converted to be Figure 2(b) by the following steps:

1. Add a new dummy node S as a source (with supply $\sum_{i=1}^{N}(\tilde{I}_i^t-I_i^t)$) or a sink (with demand $\sum_{i=1}^{N}(I_i^t-\tilde{I}_i^t)$).
2. Add $2N$ dummy arcs (j_R^t, S) and (S, j_R^t), $j=1,\dots,N$.
3. Associate each left node i_L^t with a supply $\tilde{I}_i^t$, each right node i_R^t with a supply I_i^t
4. Associate each dummy arc with a cost 1, each inventory arc with cost 0; and each repositioning arc with cost ε.



(a) (b)

Figure 2: An illustrative VRFM example

Now we can easily see the transformed graph is a minimum cost flow problem, where flows are sent from supply nodes (including the left nodes or S) to demand nodes (including the right nodes or S) via uncapacitated arcs with minimum total costs. The optimal solution to this minimum cost flow problem gives an optimal repositioning strategy.

Note that the above network transformation ignores constraints (8), the availability of voluntary riders. If, the optimal solution to the minimum cost flow problem requires to hire more than R^t voluntary riders, then we can randomly remove the extra riders from the solution without affecting the optimality. Take Figure 2 for an example, an optimal solution may reposition 4 bikes from 2 to 4, leaving 2 bikes and 1 bike shortage in site 1 and 3, respectively. However, if we can at most hire 3 voluntary riders, then we may simply lay off 1 rider from previous arrangement.

4. DATA ANALYSIS AND TESTINGS

4.1 Random Forest and Inputs for VRFM

Any real-time repositioning decision requires accurate prediction on the bike rentals in the near future. To this end, we use the random forest algorithm (Breiman, 2001) used in machine learning to construct multiple decision trees for classification of important features and regression on the rental demands. In our testings, the random forest algorithm has better prediction errors than linear regression and ARIMA.

Using the 10-month historical YouBike rental data in 2014 as the dataset, we have found that the classification errors converges after constructing more than 150 decision trees. We decide to construct 500 trees since it has almost the same errors as 3000 trees but consumes much shorter time. Based on our testings, we have selected 7 important features that affect the prediction: holiday or not, weekday or not, day (i.e., Monday, …, Sunday), current hour, number of checking out and returns in previous period, temperature, and rainfall. For each site, we construct its own random forest, which can output a predicted value on the number of checking outs ($\bar{b}_i^t$) and returns ($\bar{r}_i^t$) with the given 7 parameters.

To solve for VRFM^t, one still needs to estimate $\bar{x}_{ji}^{t^-}$, the expected number of actual voluntary riders who are currently on the way from site j and expected to arrive at site i in this period. This can be estimated from the historical data. For example, if on average it takes 1.6 periods to move from site j to i, and there are 2 and 3 riders have traveled for 0.3 and 0.9 periods at present time, then we would expect 3 riders to arrive in this period, while the other 2 riders will arrive in next period.

Therefore, with current bike inventory $\bar{I}_i^{t-1}$, target optimal inventory I_i^t, predicted number of checking outs $\bar{b}_i^t$ and returns $\bar{r}_i^t$, the expected number of actual voluntary riders $\bar{x}_{ji}^{t^-}$, and the estimated number of total available voluntary riders R^t, we can solve the VRFM^t for optimal voluntary rider flows $\tilde{x}_{ij}^t$ that we try to hire for this period. Then, after one period, we repeat the same procedures, until the end of a day.

4.2 Simulations on Repositioning Strategies

In order to test the effectiveness of our crowdsourced repositioning strategy, we use two simulation settings: (1) sampled real daily data (e.g. the rental data of 2014.03.05 or 2014.03.11), and (2) sampling from a set of mixed real (e.g. if we mix the actual 85 sunny weekdays in the period of 2014.01-05, then each rental record has 1/85 probability to be selected; We repeat these sampling for an entire day for 100 times). The first simulation uses the Random Forest algorithm to predict the rental demands, since the environment info such as temperature or rainfall are also available as inputs. Nevertheless, the second simulation can only use the historical average statistics (e.g., average checking outs/ins in each site and period) for rental demand prediction, since the random sampling would violate the data consistency for using the random forest algorithm (e.g., previous rental records may correspond to different dates, whose environment data might be very different, and thus not applicable for the Random Forest algorithm.). We use the optimal initial and ending bike inventory calculated by IIM in each period t for setting the initial and target optimal bike inventory in each period for simulation.

To simulate the crowdsourced repositioning strategy, we first assume we can always find voluntary riders whenever necessary, which gives us an estimate for R^t, the maximum number of available voluntary riders in each period t. In our testings, we find by inviting $\sum_{t=1}^{T} R^t = 4458$ daily voluntary riders the simulation would satisfy 94%~98% daily demands.

Besides the crowdsourced repositioning strategy, we also implement a truck repositioning strategy for comparison. To this end, we first conduct a K-means clustering algorithm to partition the rental sites into disjoint service zones, so that each service zone has similar amount of total net rental data (i.e. the difference between the total checking outs and returns) and the rental sites in one service zone are in close vicinity to each other. The repositioning tasks within one service zone are conducted by a unique repositioning truck. We also assign repositioning missions every 30 min, based on the real practices of YouBike (i.e., 10-15 min to move to another site in the same service zone, 10-15 min to load/unload bikes). We assume a truck can carry at most 20 bikes, and every 30 min it will select a site that requires the most loading or unloading operations. In particular, for a truck currently carrying 3 bikes at a site, there are 4 other sites in the same service zone that respectively need +5, -2, +10, -14 bikes ("+" means a site needs to add bikes, whereas "-" means a site needs to remove bikes), then the truck would go to the 4th site to move 14 bikes to the truck. Similar process is conducted every 30 min.

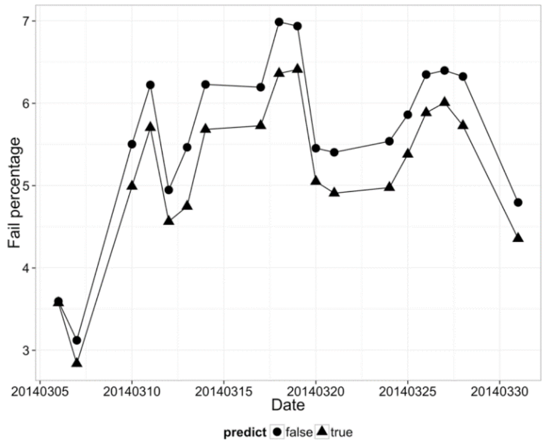


Figure 3: Comparison on the effectiveness of prediction for crowdsourced repositioning

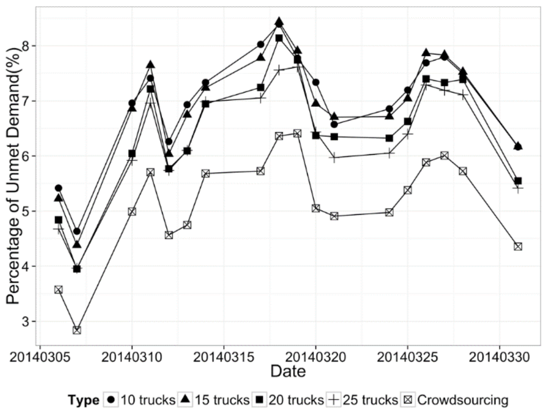


Figure 4: Comparison on the effectiveness of repositioning by crowdsourcing and different number of trucks

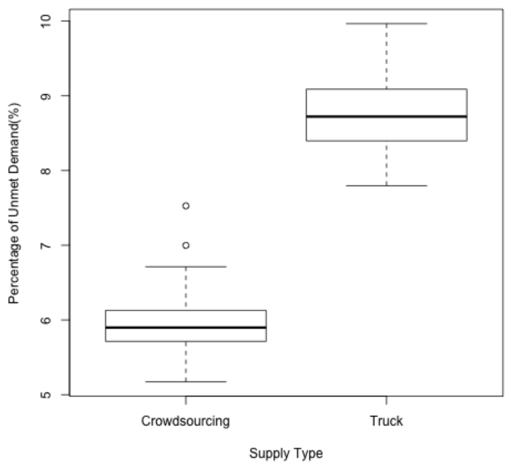


Figure 5: Comparison on the effectiveness of different repositioning strategies in 100 simulated daily rentals

Using the first simulation setting of 18 real daily rentals, Figure 3 shows that the effectiveness of prediction by the Random Forest algorithm is around 0.8%. Figure 4 shows that the crowdsourced repositioning strategy has improved up to 1-3.5% of unmet demands than repositioning by 10 to 25 trucks for the 18 actual daily rental data. These results indicate the more accurate prediction (e.g., by the Random Forest algorithm) does improve the service quality. In addition, the crowdsourcing repositioning provides much better service than trucks. Finally, Figure 5 again certificates the effectiveness of the crowdsourced repositioning is up to about 2.8% better on average than the truck repositioning.

5. CONCLUSIONS

In order to reduce the unmet rental demands in the bike sharing systems, we propose a novel crowdsourced repositioning scheme and show that it is more effective than current truck repositioning strategy by simulations. Although similar idea has been mentioned for years, our work is arguably the first one to give detailed mathematical models and numerical experiments on how to implement it, to the best of our knowledge.

We first point out the drawbacks of truck repositioning, then propose a mathematical programming model (IIM) to calculate the ideal optimal bike inventory for each site in each time period, assuming unlimited availability of voluntary riders. The calculated ideal bike inventories for each time period t are then used as a target value to reach in our second simplified linear program model (VRFM^t), which we have shown to be a minimum cost flow problem. Given the current bike inventory, estimated bikes to be checked out and returned, estimated voluntary repositioning to be completed, and estimated maximum number of available voluntary riders in the incoming period, one can solve VRFM^t for an optimal number of voluntary riders for specific OD pairs (i.e. how many voluntary riders to assign for repositioning bikes from which origin site to which destination site). To have more accurate in the estimated rental demands, we have used the Random Forest algorithm to identify important factors and parameters. To validate the effectiveness improvement in service qualities, we have conducted two simulation experiments using the 10-month real rental data collected from YouBike. The results indicate that more accurate prediction could improve the service quality up to 0.8%, and our crowdsourced repositioning strategy can improve up to 3.5% and 2.8% of service quality than the truck repositioning strategy in our two simulation experiments.

For future research, we suggest to investigate more accurate prediction models on rental demands, as well as marketing strategies in encouraging the voluntary riders.

ACKNOWLEDGMENTS

I-Lin Wang was partially supported by Sayling Wen Cultural & Educational Foundation and Service Science Society of Taiwan, as well as the Ministry of Science and Technology of Taiwan under Grant MOST 102-2221-E-006-141-MY3 and 105-2221-E-006 -182 -MY3.

REFERENCES

Bargar, A., Gupta, A., Gupta, S., and Ma, D. (2014) Interactive visual analytics for multi-city bikeshare data analysis. *The 3rd International Workshop on Urban Computing (UrbComp 2014), New York, USA.*

Benchimol, M., Benchimol, P., Chappert, B., de la Taille, A., Laroche, F., Meunier, F. and Robinet, L. (2011) Balancing the stations of a self-service bike hire system. *RAIRO – Operations Research*, **45**, 37-61.

Breiman, L. (2001) Random forests. *Machine learning*, **45**, 5-32.

Cagliero, L., Cerquitelli, T., Chiusano, S., Garza, P., and Xiao, X. (2016) Predicting critical conditions in bicycle sharing systems. *Computing*, 1-19.

Chang, L.-C. (2010) Design and Management of Urban Bike Sharing Systems. Master *Thesis, Department of Industrial and Information Management, National Cheng Kung University, Tainan 701, Taiwan.*

Chemla, D., Meunier, F. and Calvo, R. W. (2013) Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization*, **10**, 120-146.

Contardo, C., Morency, C., and Rousseau, L.-M. (2012) Balancing a dynamic public bike-sharing system. *Centre Interuniversitaire de Recherche sur les Réseaux d'Entreprise, la Logistique et le Transport (CIRRELT)*, **4**.

Froehlich, J., Neumann, J. and Oliver, N. (2009) Sensing and Predicting the Pulse of the City through Shared Bicycling. *International Joint Conference on Artificial Intelligence*, **9**, 1420-1426.

Gebhart, K., and Noland, R.B. (2014) The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, **41.6**, 1205-1225.

Hung, C.-P. (2011) Optimal Station Allocation and Dynamic Bike Repositioning Strategies for Public Bike Sharing Systems. *Master Thesis, Department of Industrial and Information Management, National Cheng Kung University, Tainan 701, Taiwan.*

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., and Banchs, R. (2010) Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, **6**, 455-466.

Kaspi, M., Raviv, T., Tzur, M., and Galili, H. (2016) Regulating vehicle sharing systems through parking reservation policies: Analysis and performance bounds. *European Journal of Operational Research*, **251**, 969-987.

Liao, M.-T. (2012) A Strategic Study on Managing Public Bike Sharing Systems by Demand Profile and Temporary Manpower Allocation. *Master Thesis, Department of Industrial and Information Management, National Cheng Kung University, Tainan 701, Taiwan.*

Montoliu, R. (2012) Discovering mobility patterns on bicycle-based public transportation system by using probabilistic topic models. *Ambient Intelligence-Software and Applications. Springer Berlin Heidelberg*, 145-153.

O'brien, O., Cheshire, J., and Batty, M. (2014) Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, **34**, 262-273.

Raviv, T., Tzur, M., and Forma, I.A. (2013) Static repositioning in a bike-sharing system: Models and solution approaches. *EURO Journal on Transportation and Logistics*, **2**, 187-229.

Rixey, R. (2013) Station-Level Forecasting of Bikesharing Ridership: Station Network Effects in Three US Systems. *Transportation Research Record: Journal of the Transportation Research Board*, **2387**, 46-55.

Sarkar, A., Lathia, N., and Mascolo, C. (2015) Comparing cities' cycling patterns using online shared bicycle maps. *Transportation*, **42**, 541-559.

Vogel, P., Greiser, T., and Mattfeld, D.C. (2011b) Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences*, **20**, 514-523.

Vogel, P. and Mattfeld, D.C. (2011a) Strategic and operational planning of bike-sharing systems by data mining–a case study. *International Conference on Computational Logistics. Springer Berlin Heidelberg*, 127-141.

Vogel, P., Saavedra, B.A.N., and Mattfeld, D.C. (2014) A hybrid metaheuristic to solve the resource allocation problem in bike sharing systems. *International Workshop on Hybrid Metaheuristics. Springer International Publishing*, 16-29.

Wang, I.-L., and Wu, T.-L. (2016) A Simulation Study on the Value and Impact of Exploiting Rental Information to the Metropolitan Bike Sharing Systems. *Technical Report, Department of Industrial and Information Management, National Cheng Kung University, Tainan 701, Taiwan*

Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., and Moscibroda, T. (2016) Mobility Modeling and Prediction in Bike-Sharing Systems. *MobiSys '16 Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, 165-178.

Yoon, J.W, Pinelli, F. and Calabrese, F. (2012) Cityride: a predictive bike sharing journey advisor. *2012 IEEE 13th International Conference on Mobile Data Management*, 306-311.