How Unbalanced are Bicycle Dynamics? Demand-Supply Shortage Detection with Spatiotemporal Tensor Factorization in station-less bike Systems

Dakshak Keerthi Chandra [*] Kunpeng Liu[$] Yanjie Fu[+,†]

Abstract

Station-less bike sharing systems are a fast growing smart transportation trend today, with more and more cities around the world implementing them. They are emerging as a novel tool that can help reduce the burden on public transport system as well as reduce traffic congestions in cities. They are also an affordable and green way to get from point A to point B for daily commuters. Station-less bike sharing systems are not constrained to bike stations like traditional bike sharing systems. By using just an app on the phone, users can rent and pick up a bike from any location most closest to them and then return it back at any arbitrary location at the end of the trip. Station-less bike sharing systems have also been instrumental in addressing and solving the commuters *last mile* problem.

However, along with all the advantages, the station-less bike sharing systems also pose new sets of challenges to be addressed and solved. Some of the primary challenges are identifying regions within the city suffering from demand shortage of bikes, supply shortage of bikes and regions with parking problems. In this paper we propose a multidimensional tensor model to address these problems. We use Mobike dataset for the city of Beijing to evaluate our model, and the experimental results show the superior performance of the proposed model.

1 Introduction

The world has witnessed a rise in popularity of station-less bike sharing system in recent years, with many cities all over the world implementing them. These station-less bike sharing systems have become especially popular in metropolitan cities like Beijing as they not only help ease the pressure of public transportation systems and help reduce traffic congestion in cities, but also provide an affordable and green way for daily commuters to travel from point A to point B. The station-less bike sharing systems are inherently different from regular bike sharing systems as the bikes are not tied down to stations. With regular bike sharing systems, the commuters needs to pick up bikes from a station closest to them and then at the end drop it off at a station nearest to the user's end location. Due to this kind of system, the regular bike sharing systems fail to address the commuter's *last mile problem* in which commuters face the problem of being stuck in a place in between their destination location and the bike station to justify the effort of picking up and dropping off the bike. However, with their ability to be station-less, the station-less bike sharing systems have been successful in addressing the commuter's *last-mile problem.* With station-less bike sharing systems, commuters don't have to face the *last mile problem* of having to pick up the bike from a station and then park the bike back in a station at the end of the trip which may or may not be close to their original starting or destination location. Station-less bike sharing system offer commuters the flexibility of picking up a bike from any location and then at the end of the trip just park the bike at a location most convenient to them. The whole system is managed using an app on the phone. Users can check for any available bikes near them using the pre-installed station-less bike app on their phone, the app displays bikes most nearest them. The user then picks a bike from the available ones for his/her trip. After the completion of the trip the user can park the bike at any point and then lock it, which signifies the end of the trip. The app records details of the trip like check-in/out time and the distance traveled based on which fee will be charged.

Along with all these advantages, due to the distinct and unique nature of station-less bike sharing systems where the user gets to pick up and drop off a bike at any location, they are facing new challenges which need to be addressed and solved. Three of the primary challenges are identifying regions within the city suffering from demand shortage of bikes, supply shortage of bikes and regions with parking problems. Since a user can pick up and drop off a station-less bike at any arbitrary location, the task of identifying regions suffering from these problems has become even more of a challenge.

[*◊+$]Department of Computer Science, Missouri S&T, Rolla, MO, USA. Email: {dkthc, kl6ph, fuyan}@mst.edu.

[†]Contact Author.

As a result busy streets of cities like Beijing are being flooded by thousands of bikes parked everywhere which in turn are adding to the already existing parking and traffic congestion problems.

Figure 1: Streets of Beijing overflowing with bikes

The main contributions of this paper are

- We propose a multidimensional model to address some of the major problems being faced by the station-less bike sharing systems.
- Our proposed framework can successfully identify problem areas within the city for different periods of time. Our model also incorporates clustering model in addition to multiple dimensions.
- In-order to deal with this kind of high amount of data as well as data sparsity we implement tensor factorization.
- We evaluate our model over Mobike datasets with more than 3.2 million trips in the city of Beijing. The experimental results verify the effectiveness of the proposed model compared with baseline models.

2 Data Description

In this section, we provide details about the Mobike and POI data sets that we have used in developing our multidimensional tensor model based on Tensor Factorization. Table 1 shows the statistics of our real-world data sets.

The Mobike trip dataset was released by Mobike in the Mobike Big Data Challenge 2017. The dataset is from the time period of May-10 to May-24 of 2017. It contains details of 3,214,096 trips along with 7 attributes associated with them. For each trip the associated attributes include details like the userid, orderid, trip start_time, trip start_location, trip end_location, etc. The initial data analysis revealed that the 84.29% of the entire dataset was made up by loyal users of Mobike who had regular routes and fixed patterns of bike usage. Out of the remaining 15.71% of the data which was made up by new/fickle users, 2.92% of the users had similar bike usage patterns and routes as the loyal users, and the remaining 12.83% of the data was generated by new/fickle customers who did not have the same bike usage patterns and routes. With these results we were able to conclude that 87.22% of the data had similar pattern. During the data analysis we considered users who rented bikes more 7 times during the entire 14 days time period as loyal users and the users who rented bikes less than 7 times were categorized as new/fickle customers. The result from the analysis drove us to the conclusion that the bike usage data had a consistent pattern and can be used to our advantage while identifying demand shortage regions, supply shortage regions and regions with parking problem.

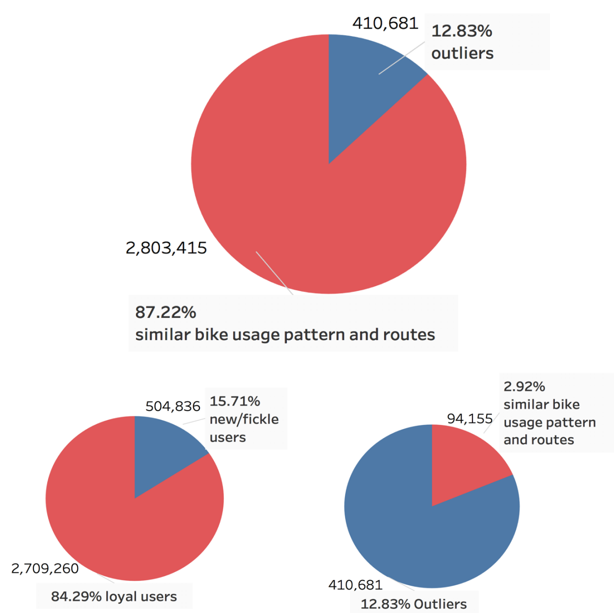

Figure 2: Data content identification

The POI data set for the city of Beijing was obtained from www.dianping.com, which is a commercial review and recommendation website. It consists of 328,668 POI's divided into 20 different categories like Hospitals, Malls, Restaurant, theaters, etc. We use the POI dataset to cross reference it with Mobike data and identify high activity regions within the city.

Table 1: Statistics of the Datasets.

Data Sources	Attributes	Statistics
Mobike Trips	Number of trip records	3,214,096
	Number of users	349,693
	Trip start time	
	Trip start location	
	Trip end location	
	Bike id	
	Time period of records	05/10/2017 - 05/24/2017
POIs	Number of POIs	328,668
	Number of POI categories	20

3 Problem Definition and Framework

In this section we provide some key definitions and also give a brief overview of the proposed framework.

3.1 Problem Definition

We initially state the below definitions to help break down the problem for better understandability. We then go on to give a proper definition of the problem as well.

Definition 1: *(Trip):* A bike trip is defined as $Trip = \{T_{id}, T_{Sloc}, T_{Eloc}T_t\}$, where T_{id} denotes the trip's unique order id, T_{Sloc} consists of the latitude and longitude point of the trip's starting location, similarly T_{Eloc} consists of latitude and longitude point of trip's ending location, and T_t denotes the trip' start time.

Definition 2: *(Bike Supply Shortage Region):* A region r is defined as a bike supply shortage region S_r at time t if the number of CheckIn bikes are smaller than the number of CheckOut bikes.

Definition 3: *(Bike Demand Shortage Region):* A region r is defined as a bike demand shortage region D_r at time t if the number of CheckIn bikes are greater than the number of CheckOut bikes.

Definition 4: *(Bike check-in Tensors):* The three dimensional bike check-in tensor can be denoted as $\mathbf{X} \in \mathbb{R}^{I\times J\times K}$. The check-in tensor contains data about the number of bikes checked in at a particular region for each day and hour.

Definition 5: *(Bike check-out Tensors):* The three dimensional bike checkOut tensor can be denoted as $\mathbf{Y} \in \mathbb{R}^{I\times J\times K}$. The check-out tensor contains data about the number of bikes checked out from a particular region for each day and hour.

Definition 6: *(Bike Parking Problem Region):* A region r is defined as a bike parking problem region P_r at time t if the total number of bikes present in region$r$ at time t is greater than the predefined threshold P_{th}.

Problem Definition. Given a dataset consisting of bike trips along with their origin location, destination location, trip start time, order id, user id, and POI's of a city, our objective is to identify regions of the city with bike demand/supply shortage and parking problems.

3.2 Brief Overview of the Framework

Figure 4 shows the framework of our proposed model. It consists of 4 major steps: preliminary data analysis and clustering model, tensor construction and factorization, identification of problem areas and optimization.

During the preliminary data analysis stage we leverage the POI data for the city of Beijing along with the bike check-in/out data from the Mobike dataset and cross-reference both of them to identify regions of the city with high activity. Then we construct 2 three dimensional tensors, one for bike check-in data and the other for bike check-out data. Here, regions, days and hours make up the 3 dimensions of the tensors. This enables us to capture the check-in/out data of every region for each hour of each day. In the identification of problem areas step we use the two constructed tensors to identify total number of bikes present in each region at specific times. We then use the resulting tensor to identify problem areas. Since we are dealing with multidimensional data the parameters required to be stored in a tensor can increase exponentially and also to deal with data sparsity problem we propose using tensor factorization method.

4 Constructing Multidimensional Model

We use context information to construct our multidimensional model.

4.1 Preliminary Data Analysis and Clustering

We use the POI dataset for the city of Beijing to first identify regions with highest number of POI's in the ci-

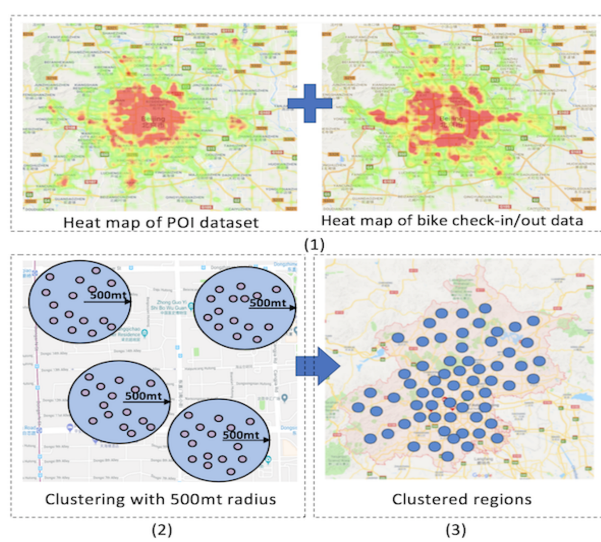

Figure 3: Clustering model

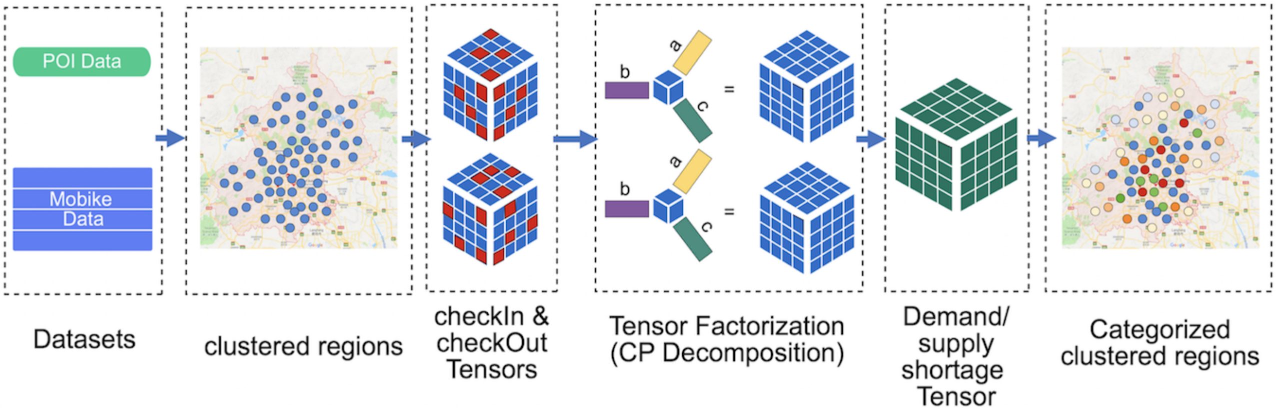

Figure 4: Framework of the proposed model

ty. The regions with higher number of POI's are considered as high human activity regions. We then also analyze the bike check-in and check-out data to identify regions which have high bike activity. We then cross-reference the high human activity regions with high bike activity regions which enables us to identify regions with highest activity in the city of Beijing in terms of POI's and bike activity. After identifying the highest activity regions we implement our clustering model on them to group the locations within the regions into clusters, with each cluster having a radius of 500mt radius. Each cluster contains n number of both POI locations and bike check-in, check out locations. The whole flow of the process of data analysis and clustering can be summed up in 3 steps as shown in Figure 3.

4.2 Tensor Construction

Based on the clustered model we take into account clustered regions which can also be considered as virtual stations to construct 2 tensors. One for bike check-in data and the other for bike check-out data. For both

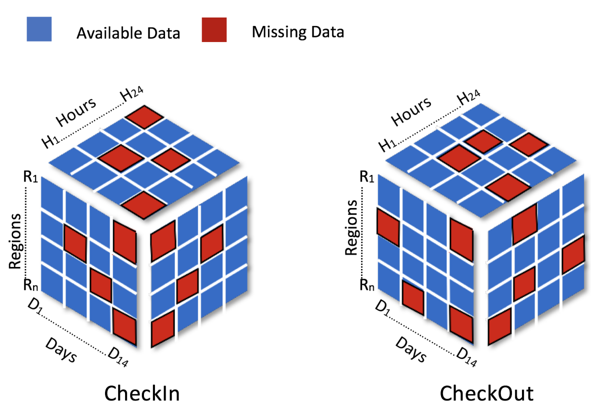

Figure 5: check-in and check-out tensors

check-in and check-out tensors we represent each clustered region $i_1, i_2....i_n$ as the first dimension. We represent days $j_1, j_2....j_n$ and hours in a day $k_1, k_2....k_n$ as second and third dimensions. Check-in tensor $\mathbf{X}$ is represented as $\mathbf{X} \in \mathbb{R}^{I \times J \times K}$ and check-out tensor $\mathbf{Y}$ is represented as $\mathbf{Y} \in \mathbb{R}^{I \times J \times K}$.

4.3 Tensor Factorization

In-order to account for the missing data and to compress the data we implement tensor factorization using CP decomposition method[11]. We use the SiLRTC algorithm[15] in-order to detect missing values in a tensor. With this method we are not only able to retrieve the missing data but also will be able to store the data in its compressed form which can be retrieved later on for any future operations. The CP decomposition method factorizes a tensor into sum of a finite number of rank-one tensors. For example, given a third-order tensor $\mathbf{M} \in \mathbb{R}^{I \times J \times K}$ we can write it as

$$\mathbf{M} \approx \sum_{r=1}^{R} a_r \circ b_r \circ c_r \tag{4.1}$$

where R is a positive integer and $a_r \in \mathbb{R}^I$, $b_r \in \mathbb{R}^J$, and $c_r \in \mathbb{R}^K$. This can also be written elementwise as

$$m_{ijk} \approx \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} \tag{4.2}$$

$$for\ i = 1, ..., I, j = 1, ..., J, k = 1, ..., K.$$

4.4 Fusing tensors for problem area identification

By fusing the two completed tensors we can identify demand/supply shortage regions and regions suffering from parking problems. By fixing the two dimensions of the tensor and performing matrix subtraction operation we are able to identify regions with demand/supply shortage problems. Now, in-order to identify regions suffering from parking problems we analyze the data based on the trip start time.

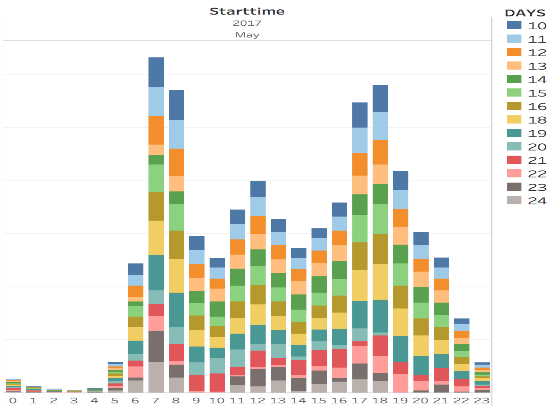

Figure 6: Bike usage pattern

By doing so, we were able to identify a pattern between bike usage and particular hours in a day. Figure 6 shows the pattern. We observed that the bike usage is high during 7AM and 8AM in the morning, and during 5PM, 6PM and 7PM in the evening. This is probably because of the number of people going to work and coming back from work after a night shift in the morning and people returning back home after work or people going to watch a movie or eat at a restaurant after getting off from work in the evening. This can be true not only for people who are using bikes to commute but also for people who are using taxi's, bus's and other means of transportation. We also observe a slight rise in bike usage from 11AM to 1PM. This might be due to people going out for lunch during break and people coming in for their afternoon shift at work. It is safe to assume that during these hours the probability of a region suffering from traffic congestion as well as parking problem is at its peak. We detect regions with parking problems by initially setting a parking threshold R_{pt}. We set a parking threshold while taking into consideration the fact that we're only analyzing the dataset of mobike and that there are multiple companies like mobike offering similar services whose bikes may also be present in the region being analyzed. We calculate parking threshold for individual regions depending on the growth rate of bikes parked in those regions. Given a region r and an initial parking threshold Pi_{th} with number of bikes parked in region r at hour 0 denoted as r_ih_0 and number of bikes parked in region r at hour 23 denoted as r_ih_{23}. We can then determine parking thresholds of individual regions.

(4.3)

$$\begin{cases} R_{1pt} = Pi_{th} + \{[(r_1h_0/r_1h_{23})^{1/24} - 1] \times Pi_{th}\} \\ R_{2pt} = Pi_{th} + \{[(r_2h_0/r_2h_{23})^{1/24} - 1] \times Pi_{th}\} \\ R_{3pt} = Pi_{th} + \{[(r_3h_0/r_3h_{23})^{1/24} - 1] \times Pi_{th}\} \\ R_{4pt} = Pi_{th} + \{[(r_4h_0/r_4h_{23})^{1/24} - 1] \times Pi_{th}\} \\ . \\ . \\ . \\ . \\ R_{npt} = Pi_{th} + \{[(r_nh_0/r_nh_{23})^{1/24} - 1] \times Pi_{th}\} \end{cases}$$

generic form of the equation can be written as,

$$R_{ipt} = Pi_{th} + \{[(r_ih_0/r_ih_{23})^{1/24} - 1] \times Pi_{th}\}$$

The parking threshold R_{ipt} value differs for different regions. The region r is considered as a region suffering from parking problem P_r at time t if the total number of bikes present in that region are greater than the predefined threshold. Algorithm 1 gives the details of the procedure.

Algorithm 1 Supply, Demand and Parking problem regions detection

Input: CheckIn tensor $\mathbf{X}$, checkOut tensor $\mathbf{Y}$ and set of parking thresholds p for different categories
Output: Identification of demand shortage regions D_r, supply shortage regions S_r and parking problem regions P_r

Initialize tensors $\mathbf{X}$ and $\mathbf{Y}$ after fixing any two dimensionsd_0, d_1 provided both fixed dimensions d_0 and d_1 are same for the two tensors.

```
if d_1, d_2 of X == d_1, d_2 of Y then
    Z = X − Y
    if Z_{i,j,k} > 0 then
        Z_{i,j,k} = D_r
    end
    else if Z_{i,j,k} < 0 then
        Z_{i,j,k} = S_r
    end
    for p_1, p_2, p_3...p_n do
        for f in [1, k] do
            if r(Z_{i,j,k}) > p then
                r(Z_{i,j,k}) = p_r
            end
        end
    end
end
```

With this we are also able to determine that parking problem of a region is independent of supply shortage and demand shortage factor of a region. It is solely dependent on the predefined threshold and individual hours of a day.

5 Model Evaluation

We evaluate our model for two criteria.(i) We implement tensor factorization on different ranks to test which rank gives the optimal accuracy while keeping the number of parameters to be stored at an acceptable limit when the tensor is decompressed. (ii) We test our model against different methods to examine the accuracy of the predicted missing values recovered by implementing tensor factorization. Based on this we determine its effectiveness.

Root Mean Square Error (RMSE)[2] measures how much error there is between two data sets. In other words, it compares a predicted value and an observed or known value. The formula for calculating RMSE is as below.

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(y_j - \hat{y}_j\right)^2}$$

It quantifies how different a set of values are. The smaller an RMSE value, the closer predicted and observed values are.

In-order to find the optimal rank we factorize our tensors on different ranks and then check the RMSE error between the original tensor and the recovered tensor. We compare our results with matrix factorization method.

Table 2: RMSE values for different ranks

	Matrix Factorization		Tensor Factorization	
Rank	RMSE	Parameters	RMSE	Parameters
1	5.04315	1286	3.16346	988
2	3.54652	2572	2.75363	1976
3	3.47987	3858	2.50851	2964
4	2.61652	5144	2.22143	3952
5	2.57374	6430	2.12987	4940

In the experiment we implement tensor factorization as well as matrix factorization on the data set while keeping the rank same for both methods. The experimental results lead us to conclude that tensor factorization not only provides better accuracy while reconstructing the original tensor from the factorized tensor, but can also keep the number of parameters to be stored at minimal when compared with matrix factorization.

For the second criteria we evaluate our model against other models to determine its effectiveness. We test our model against Matrix Factorization and Linear Regression for different percentage of missing data.

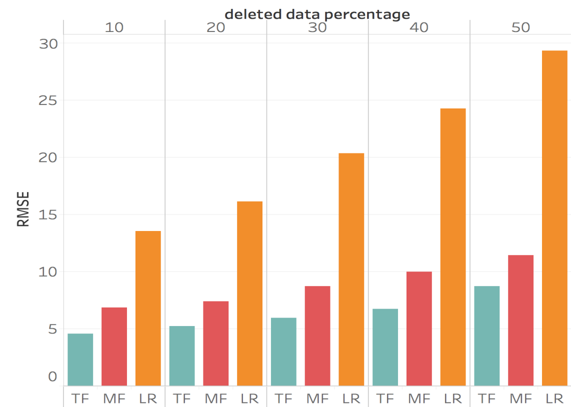

Figure 7: RMSE values of TF, MF and LR for different missing data percentage

Matrix Factorization is one of the most widely used method to predict missing values and also to compress the data. Models based on Matrix Factorization have received greater exposure, mainly as an unsupervised learning method for latent variable decomposition and dimensionality reduction [12][5]. It is most similar to tensor factorization method. In MF method, a matrix V is factorized into two matrices W and H and the missing values are approximated numerically.

Linear Regression is also one of the most widely used methods to predict missing values. Its broad appeal and usefulness results from conceptually logic process of using an equation to express the relationship between a variable of interest and a set of related predictor variables. It uses the relationship between scalar dependent variables and one or more explanatory variables to predict missing values[10].

The main aim of this evaluation is to check which model can better predict the missing data. We measure accuracy using RMSE method. In the first case where there is only 10% of the data is missing, the RMSE value is considerably low, but as we increase the percentage of deleted data the RMSE value also increases. In the final test case where we test against 50% of the missing data, we can see that RMSE value increases considerably more. In all the test cases, we observe that RMSE value of tensor factorization manges to be lower when compared with values of other methods and based on this we conclude that our model performs better than the other two methods. Figure 7 shows the result of our experiment.

6 Related Work

Bike sharing systems have attracted a lot of research interest from the beginning. A number of recent researches have tried to address the problems posed by stationless bike sharing systems. The following are the most current research works that are closely related to our work. In 2013 Chemla et al. [3] proposed the static re-balancing problem paper, which deals with balancing Demand/supply of bikes. The paper address the problem of redistribution of bikes to different location. It presents efficient algorithms for solving instances of reasonable size, and contains several theoretical results related to this problem. Faghih-Imani et al. studies the decision process involved in identifying destination locations after picking up a bicycle from a shared-bike station, in the form of a multinomial logit model[6], The paper by Chen et al. [4] propose a dynamic cluster based model to predict over demand of bikes taking into account the common contextual factors opportunistic contextual factors that affect the bike usage pattern. Work by Yang et al. propose a spatio-temporal bicycle mobility model based on historical bike-sharing data, and devise a traffic prediction mechanism on a per-station basis[23]. Zhang et al. introduce a new trip destination prediction and trip duration inference model on the basis of analyzing individuals bike usage behaviors on traditional bike sharing systems[24]. Singla et al. proposed a incetivizing approach for balancing bike sharing systems. The authors propose to engage the users themselves to solve the imbalance problem in bike sharing systems by providing them incentives to ride bikes from station suffering from demand shortage of bikes[21]. Liu et al [16] proposed a model for bike re-balancing and data optimization. Meng et al. in 2011 wrote a paper in which they proposes a complete methodology for introducing bike-sharing systems. The proposed methodology takes into account potential demand for bicycle use and the willingness to pay of future users for faster journey times, and also introduces a location model for fixing the bicycle pick-up and drop-off stations made with the help of a geographical information system[17]. Lin and Yang in 2011 study the strategic planning of public bicycle sharing systems while considering the interests of both users and investors, the proposed model attempts to determine the number and locations of bike stations[14]. Caggiani et al. proposed a flexible fuzzy decision support system for redistribution process in traditional bicycle sharing systems is presented with the main aim to minimize the redistribution costs for bike-sharing companies, determining the optimal bikes repositioning flows, distribution patterns and time intervals between relocation operations, with the objective of a high level for users satisfaction[1]. Froehlich et al. proposed a model which adopted a Bayesian network to predict station status based on the current time and current available dock number[7]. Kaltenbrunner et al. [8] proposed a short term prediction of the number of available bikes in stations via the analysis of cyclic mobility patterns. It detects temporal and geographic mobility patterns which are applied to predict the number of available bikes for any station. The predictions are used to improve the bicycle program. Li et al. proposed a hybrid and hierarchical pre- diction model to predict the number of bikes that will be rent from/returned to each station cluster in the early future[13]. Espegren et al. in 2015 wrote a a paper that considers the static bicycle repositioning problem (SBRP), which deals with optimally re-balancing bike sharing systems (BSS) overnight by using service vehicles to move bikes from full stations to empty stations. A new and improved mathematical formulation for the SBRP is proposed[5].

Our work is also very closely related to Tensor Factorization. Rendle et al. proposed Factorization machines that combines the advantages of Support Vector Machines (SVM) with factorization models[19]. In 2010 Xiong proposed a temporal collaborative filtering with bayesian probabilistic model using tensor factorization[22]. The work by Oentaryo develops a Hierarchical Importance-aware Factorization Machine (HIFM), which provides an effective generic latent factor framework that incorporates importance weights and hierarchical learning[18]. The work by Karatzoglou introduces a Collaborative Filtering method based on Tensor Factorization (TF), with types of context considered as additional dimensions in the representation of the data as a tensor[9].The work by presents the factorization model PITF (Pairwise Interaction Tensor Factorization) which is a special case of the TD model with linear runtime both for learning and prediction[20].

7 Conclusion

In this paper, we propose a model which combines a cluster model and a multidimensional tensor based model to address some of the major problems faced by station-less bike sharing systems. We use our model to predict and detect areas with demand/supply shortage of bikes and areas with parking problems. In-order to achieve this, we combine Mobike dataset for the city of Beijing and the POI dataset to build our model. We first use the POI dataset along with bike check-in and check-out data to identify regions with high activity. We then implement clustering model to divide these high activity regions into virtual stations. Then we construct our multidimensional tensor based model. We start construction of our model by building two separate tensors, one for check-in data and the other for check-out data of bikes, then we implement tensor factorization to predict any missing values within the data and to compress data. Final steps consists of fusing the two tensors which enables us to successfully identify regions with demand/supply shortage of bikes and parking problem. The unique perspective of our model is due to the combination of clustering model and the multidimensional tensor model. Experimental results show that effectiveness of our model.

References

[1] L. Caggiani and M. Ottomanelli. A modular soft computing based method for vehicles repositioning in bike-sharing systems. *Procedia-Social and Behavioral Sciences*, 54:675–684, 2012.

[2] T. Chai and R. R. Draxler. Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature. *Geoscientific model development*, 7(3):1247–1250, 2014.

[3] D. Chemla, F. Meunier, and R. W. Calvo. Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization*, 10(2):120–146, 2013.

[4] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.-M.-T. Nguyen, and J. Jakubowicz. Dynamic cluster-based over-demand prediction in bike sharing systems. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 841–852. ACM, 2016.

[5] H. M. Espegren, J. Kristianslund, H. Andersson, and K. Fagerholt. The static bicycle repositioning problem-literature survey and new formulation. In *International Conference on Computational Logistics*, pages 337–351. Springer, 2016.

[6] A. Faghih-Imani and N. Eluru. Analysing bicycle-sharing system user destination choice preferences: Chicagos divvy system. *Journal of Transport Geography*, 44:53–64, 2015.

[7] J. Froehlich, J. Neumann, N. Oliver, et al. Sensing and predicting the pulse of the city through shared bicycling. In *IJCAI*, volume 9, pages 1420–1426, 2009.

[8] 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*, 6(4):455–466, 2010.

[9] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 79–86. ACM, 2010.

[10] G. T. Knofczynski and D. Mundfrom. Sample sizes when using multiple linear regression for prediction. *Educational and psychological measurement*, 68(3):431–442, 2008.

[11] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[12] D. kumar Bokde, S. Girase, and D. Mukhopadhyay. Role of matrix factorization model in collaborative filtering algorithm: A survey. *CoRR, abs/1503.07475*, 2015.

[13] Y. Li, Y. Zheng, H. Zhang, and L. Chen. Traffic prediction in a bike-sharing system. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 33. ACM, 2015.

[14] 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*, 47(2):284–294, 2011.

[15] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–220, 2013.

[16] J. Liu, L. Sun, W. Chen, and H. Xiong. Rebalancing bike sharing systems: A multi-source data smart optimization. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1005–1014. ACM, 2016.

[17] L. D. MEng et al. Implementing bike-sharing systems. *Proceedings of the Institution of Civil Engineers*, 164(2):89, 2011.

[18] R. J. Oentaryo, E.-P. Lim, J.-W. Low, D. Lo, and M. Finegold. Predicting response in mobile advertising with hierarchical importance-aware factorization machine. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 123–132. ACM, 2014.

[19] S. Rendle. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3):57, 2012.

[20] S. Rendle and L. Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 81–90. ACM, 2010.

[21] A. Singla, M. Santoni, G. Bartók, P. Mukerji, M. Meenen, and A. Krause. Incentivizing users for balancing bike sharing systems. In *AAAI*, pages 723–729, 2015.

[22] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 211–222. SIAM, 2010.

[23] Z. Yang, J. Hu, Y. Shu, P. Cheng, J. Chen, and T. Moscibroda. Mobility modeling and prediction in bike-sharing systems. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pages 165–178. ACM, 2016.

[24] J. Zhang, X. Pan, M. Li, and S. Y. Philip. Bicycle-sharing system analysis and trip prediction. In *Mobile Data Management (MDM), 2016 17th IEEE International Conference on*, volume 1, pages 174–179. IEEE, 2016.