Mining Spatial and Temporal Movement Patterns of Passengers on Bus Networks

Chunjie Zhou*, Pengfei Dai†, Fusheng Wang‡, Renpu Li*
*School of Software, Ludong University, Shandong, China, Email: lucyzcj@gmail.com
†Institute of Network and Technology, ICT(Yantai), Shandong, China, Email: daipf@int-yt.com
‡Department of Biomedical Informatics, Stony Brook University, New York, USA, Email: fusheng.wang@stonybrook.edu

***Abstract*—The analysis of human behavior is the basis of understanding many social phenomena. Accurate and reliable human movement pattern mining can lead to instructive insight to transport management, urban planning and location-based services (LBS). As one of the most widely used forms of transportation, buses can tell a lot of stories about people, including passenger demands, areas people are interested in crossing each day, and their travel patterns. Based on a large database from a real bus system, this paper aims to mine spatial and temporal movement patterns of passengers: evaluating traveling time of passengers, predicting number of passengers to estimate passenger demand and the crowdedness in the bus, and identifying attractive areas for passengers. There are major challenges for mining human movement patterns on bus networks: inhomogeneous, seasonal bursty periods and periodicities. In this paper, we take a Poisson process approach to model and evaluate traveling time of passengers, which can reflect the time features of individuals and activity cycles among areas. To overcome the challenges, we propose three prediction models and further take a data stream ensemble framework to predict the number of passengers. To obtain meaningful patterns of attractive areas, we provide a hierarchical clustering based approach to group spatiotemporally similar pick-up and drop-off points, as people's interests to these areas vary significantly on time, days and seasons. Our performance study based on a real dataset of five months' bus data demonstrates that our approach is quite effective: among 86,411 passenger demands on bus services, more than 78% of them are accurately forecasted.**

Keywords–movement pattern; traveling time; attractive areas; passenger demand; hierarchical clustering.

I. INTRODUCTION

Human activity patterns have received a certain amount of attention in recent studies. Analyzing human activity data to obtain structural information of humans has become an important means of studying social systems. Analyzing of individual banknotes and civil aviation traffic are notable studies in recent years. These studies have shown that individuals follow simple and reproducible patterns of mobility in several different manners [1]. Mining human movement patterns can help us to understand urban form and travel composition which can be used to support urban planning in terms of facility location and site selection. Meanwhile, it also helps us to explore passenger travel demand in different areas and in different time periods from transport management perspective, which is very useful for bus management. From LBS providers' perspective, the knowledge of passenger movement and behavior pattern can help to provide better tailored LBS, such as point of interests (POI) recommendation for a given time within appropriate scope.

As one of the most widely used mode of transport, bus can tell a lot of stories. It can tell not only road network traffic condition, but also areas people are interested in crossing in a day and their related travel patterns, such as traffic demand and their movement. Conventional bus information analysis tends to focus on road network travel time and average speed estimation. Indeed, as bus services are required by the same individual during his/her daily routine, it offers a proxy to capture individual human movement patterns [2]. This provides a unique opportunity for us to take advantage of bus information to discover spatial and temporal movement patterns of passengers, a major focus of this paper.

There are three major goals of this paper. 1) Evaluate traveling time length of passengers. In order to get the time features of individuals, we follow the Poisson process to evaluate the traveling time of passengers. 2) Predict the number of passengers to estimate passenger demand and congestion degree. When traveling with buses, travelers care about not only the waiting time, but also the crowdedness in the bus. Overcrowded bus may drive away the anxious travelers and make them reluctant to take buses. We propose a novel methodology to produce online predictions on the passenger demand using time series forecasting techniques. 3) Identify attractive areas. Attractive areas are places that people often visit, for instance, hot shopping and leisure places or living and working areas based on their level of attractiveness (LoA). We take bus pick-up and drop-off points as the focus in this study, because they can convey rich information on identifying attractive areas and associated movement patterns. An area's LoA can be determined by travelers' visiting frequency, which can be measured by the number and density of bus pick-up and drop-off points.

Accurate, real-time and reliable human movement pattern mining is the basis of understanding many social phenomena. However, due to a number of stochastic variables, we need to face the following three major challenges: 1) *inhomogeneous*. A periodicity in time on a daily basis that reflects the patterns of the underlying human activity, making the data appear non-homogeneous. 2) *seasonal bursty periods*. The movement patterns of passengers can be often messed by seasonal bursty periods of expected events such as highly crowded holiday events, weather changes, and so on. 3) *other periodicities*. The passengers' demands and attractive areas vary significantly at different time of a day, different day of a week, or even different seasons.

Aiming to address these challenges, in this paper, we present methods to mine spatial and temporal movement patterns of passengers on a bus network. To predict passenger demands, we develop a unique predictive model by adapting time series forecasting techniques to our problem. To discover attractive sites, we develop a hierarchical clustering based approach to group similar pick-up and drop-off points since people's interests to these areas varies through time of the day, day of the week, even season of the year.

In our work, we have conducted a real study using a

 ISBN: 978-1-61208-383-4

dataset obtained by a large-sized bus network containing a total of 416 bus stops and 1,326 vehicles running in the City of Yantai, China. Our test-bed is a computational stream simulation running offline. The data from the first 16 weeks were used as training set and the data from the last 6 weeks were used as input for our stream-type test-bed, i.e., simulating the movement patterns that would arrive continuously in a stream. Our experiments demonstrated promising results of our approach: our model can accurately predict more than 78% of actual passenger demands for bus services.

Contributions. The major contributions of the paper are summarized below.

- We take a Poisson process approach to model and evaluate traveling time of passengers, which can reflect the time features of individuals and activity cycles among this area.
- We propose three distinct prediction models and a well-known date stream ensemble framework to predict the number of passengers to estimate passenger demand and congestion degree. These three models can gradually solve the challenges of inhomogeneous, seasonal bursty periods and periodicities.
- We provide a hierarchical clustering based approach to identify attractive areas by grouping similar pick-up and drop-off points, since people's interests to these areas varies significantly at different time of a day, different day of a week, or even different seasons.
- Our comprehensive comparative performance study based on real datasets demonstrates the effectiveness of our methods.

Paper Organization. The rest of the paper is organized as follows. Section II summarizes the related work. Section III introduces the time features of passengers. Section IV proposes three distinct prediction models and a well-known data stream ensemble framework to predict the number of passengers. In Section V we introduce our study method to identify attractive areas. Section VI describes the experiments based on a real-world scenario, including the evaluation metrics of our model, the experimental setup, and the experiment results. Section VII concludes the paper and describes the future work.

II. Related Work

Some studies [3] utilized data mining approaches, such as clustering to extract meaningful information by analyzing trajectory stops, moves and their sequences. Work in [4] attempted to address the problem by introducing semantic modeling process into trajectory points so that meaningful patterns can be extracted from trajectories, as background geographic information is of fundamental importance for both traffic-oriented and user scenario-oriented analysis. Otherwise, patterns may be incomplete due to the missing of POI [3].

Semantic model approach [5] combines background geographic information together with trajectory location (x, y), and the concepts of stops and moves [6] are often used to facilitate discovering and modeling trajectory patterns. However, semantic modeling approach is only applicable when trajectory points can be matched with background geographic places precisely. Since trajectories are often matched onto road centerlines, while POIs are distributed along road links and stored as point object in map database, this approach may not be applicable under many circumstances unless POIs are represented by polygons and trajectory points are dense enough to surround these POIs.

To generate meaningful patterns, various types of clustering techniques have been popularly used. Trajectory point density, frequency, and stay time are the most frequently used factors to assist information extraction. For example, Alvares et al. [5] extracted moving patterns which were assumed to follow Markov chain between general type of stops, such as hotels, airports and tourist places. It also assumed each stop was located at a POI, and used travel frequency to judge their "importance". Palma et al. [3] identified stops by using a density-based clustering algorithm and introducing a "minimal stop durations" which takes into account of the average periodicity of the trajectory time points. Li et al. [4] further considered users' travel experiences, and discovered the association among these points, for instance, classical travel sequence. Verhein and Chawle [7] defined spatio-temporal association rules and related concepts and found patterns using pruning properties based on synthetic dataset.

Different from the above studies which are mostly based on limited amount of personal trajectories, our study uses large amount of bus information data to explore time-dependent attractive areas and movement patterns. Such patterns can be represented by high traffic demand areas and passenger movement among them. This is more complex in terms of wider geographic coverage and diverse individual trip purposes. Since bus pick-up and drop-off points represent traffic demand, clustering these points is a feasible approach to discover areas with high travel demand. During this process, travel interactions among the clusters and other information are obtained. The detailed approach will be illustrated in the following sections.

III. Traveling Time of Passengers

As one of the main parts of mining spatial and temporal movement patterns of passengers, we first put all the buses as a whole to study the traveling time of passengers. Based on this result, we may roughly get the time features of individuals and activity cycles among this area. We follow the Poisson process to predict the traveling time and assume that at a given interval time t, the probability of n-incident occurred is:

$$P(n,q) = e^{-qt}\frac{(qt)^n}{n!} \tag{1}$$

Here, q represents the probability of an event occurrence. Based on (1), we can derive that the interval distribution of two consecutive events is:

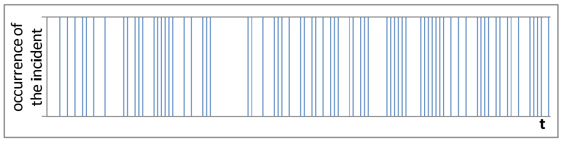
$$P(\tau) = qe^{-q\tau} \tag{2}$$

Here, τ means the traveling time of two consecutive events. To test this hypothesis, we analyze the data of traveling time and find that the distribution of traveling time over all buses is well approximated by a Power-law:

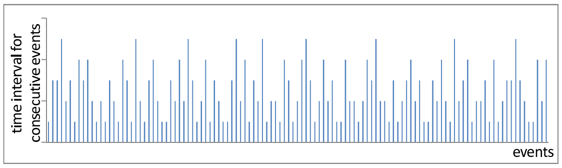
$$P(\tau)\infty\tau^{-\alpha} \tag{3}$$

Here, the exponent is between 2.20 and 2.90. We also find that the traveling time of passengers are mainly in the range of 15-30 minutes, and only a small portion of intervals are longer

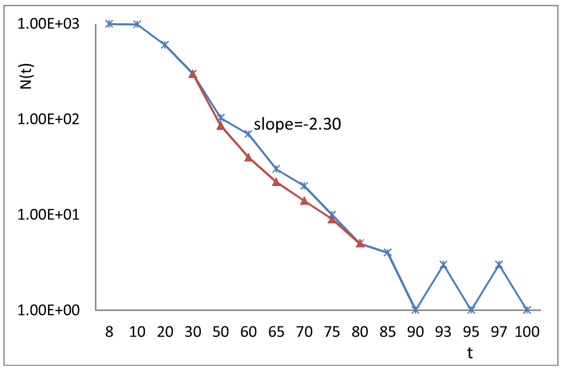
 ISBN: 978-1-61208-383-4



(a) a sequence of events.



(b) the interval time of successive events.



(c) the distribution of data points.

Figure 1. The distribution of the traveling time.

than an hour. The results of traveling time analysis are shown in Figure 1.

Figure 1 shows the traveling time distribution of passengers. Figure 1(a) shows a sequence of events. Each vertical line means an occurrence of the incident at that time. Figure 1 (b) shows the interval time of 200 successive events, and the length of the vertical line corresponds to the time interval for two consecutive events. Figure 1(c) shows the distribution of part of the data points in (a) and (b) diagrams in double logarithmic coordinates.

IV. The Passenger Demand Prediction

Our goal is to predict the number of passenger demands of a certain bus b at the bus stop s at instant t. To achieve this, we have studied three distinct prediction models, and a data stream ensemble framework.

A. Time Varying Poisson Model

The passenger demand on bus services normally follows a periodic pattern on a daily basis.

Here, we assume the probability $P(n)$ of n buses appearing in a determined time period follows a Poisson distribution. We can define it using the following equation:

$$P(n;\lambda) = \frac{e^{-\lambda}\lambda^n}{n!}, \tag{4}$$

Here, λ represents the rate (averaged number of passenger demands on bus services) in a fixed time period. However, in this specific problem, the rate λ is not constant but time-variant. Thus, we adapt it as a function of time, i.e., $\lambda(t)$, and transform the Poisson distribution into a nonhomogeneous one. Let $\lambda(t)$ be defined as follows:

$$\lambda(t) = \lambda_0 \delta_{d(t)} \eta_{d(t),h(t)}, \tag{5}$$

Here, $d(t)$ represents the weekday $\{1 = Sunday, 2 = Monday, ...\}$; $h(t)$ is the period in which time t falls in.

The model requires the validity of both equations

$$\sum_{i=1}^{7} \delta_i = 7, \tag{6}$$

$$\sum_{t=1}^{T} \delta_{t,i} = T \quad \forall t, \tag{7}$$

where T is the number of time spans in a day. To ease the interpretation of these equations, we can define the remaining symbols as follows:

- λ_0 is the average rate (i.e., expected rate) of the Poisson process over a full week;
- δ_i is the relative change for the day i (Saturdays have lower day rates than Tuesdays);
- $\eta_{j,i}$ is the relative change for the period i on the day j (the peak hours);
- $\lambda(t)$ is a discrete function representing the expected distribution of passenger demands on bus services over time for a bus stop of interest s.

B. Weighted Time Varying Poisson Model

The model above can predict the time-dependent average number of passenger demands on bus services. However, it is not guaranteed that every bus stop has highly regular passenger demands: indeed, the demands in many stops can be often messed by seasonal bursty periods of expected events such as highly crowded holiday events, weather changes, and so on.

To tackle this special seasonal issue, we propose a weighted average model based on the above presented approach. Our goal is to increase the relevance of the demand pattern observed in the last week comparing to the patterns observed several weeks ago.

Here, the weight set w is calculated using a well known time series approach – the Exponential Smoothing approach [8]. We define w as follows:

$$w = \alpha * \{1, (1-\alpha), (1-\alpha)^2, ..., (1-\alpha)^{\lambda-1}\}, \tag{8}$$

Here, λ is the number of historical periods considered in the initial average, α is the smoothing factor (i.e., a user defined parameter) and $0 < \alpha < 1$.

C. Autoregressive Integrated Moving Average Model

The last two models assume the existence of a regular (seasonal or not) periodicity in passenger demands on bus services. However, there are other periodicities, for example, the number of passenger demands in one bus stop of a certain bus in regular workdays during a certain period is highly similar. Moreover, the number of passenger demands on bus services in the morning and in the evening of the same day is also very similar.

We explore the Autoregressive Integrated Moving Average Model (ARIMA) [10]. In this model, the future value of a variable is assumed to be a linear function of several past observations and random errors. We can formulate the underlying process that generates the time series (passenger demands on bus services over time for a given bus stop s) as:

$$\begin{aligned} R_{s,t} = \theta_0 + \phi_1 X_{s,t-1} + \phi_2 X_{s,t-2} + ... + \phi_p X_{s,t-p} \\ + \varepsilon_{s,t} - \theta_1 X_{s,t-1} - \theta_2 X_{s,t-2} - ... - \theta_q X_{s,t-q} \end{aligned} \tag{9}$$

Here, $R_{s,t}$ and $\varepsilon_{s,t}$ are the predicted value and the random error at time period t respectively; $\phi_l (l = 1, 2, ..., p)$ and $\theta_m (m = 0, 1, 2, ..., q)$ are the model weights; p and q are positive integers often referred as the orders of the model. Orders and weights can be inferred from the historical time series using both the autocorrelation and partial autocorrelation functions.

D. Sliding Window Ensemble Framework

We have proposed three distinct predictive models to learn from long, medium and short-term historical data. Then, how can we combine them all to further improve our prediction?

Let $M = \{M_1, M_2, ..., M_z\}$ be a set of z models of interest to model a given time series and $M_t = \{M_{1t}, M_{2t}, ..., M_{zt}\}$ be the set of forecasted values to the next period on the interval t by those models. The ensemble forecast E_t is obtained as

$$E_t = \sum_{i=1}^{z} \frac{M_{it}}{\beta}, \beta = \sum_{i=1}^{z} \rho_{iH} \tag{10}$$

where ρ_{iH} is the forecasting accuracy obtained for the model M_i in the periods contained in the time window $[t - H, t]$. H is a user-defined parameter to define the window size. As the information is arriving in a continuous manner for the next periods $\{t, t+1, t+2, ...\}$, the window will also slide to determine how are the models performing in the last H periods. To estimate such accuracy, we take a time series forecasting error metric: the Symmetric Mean Percentage Error (sMAPE) [11].

V. Discovery of Attractive Areas of Passengers

Since each pick-up and drop-off stop can only represent the approximate location where travel demand generates from, it is more proper to represent LoA using a polygon instead of a point. Clustering is a feasible method to eliminate noises in discovering the patterns while defining the borders of attractive areas. After pick-up and drop-off points are clustered, passenger movement pattern can be identified more clearly with more information, such as flow interaction and average travel distance.

Since the number of clusters cannot be known beforehand, we take a hierarchical or agglomerative clustering algorithm instead of partition clustering. An example of partition clustering is k-mean, which requires a pre-knowledge of cluster number and the shapes of all clusters have to be convex. We follow the notion that the distance $dist(i, j)$ between two points (i, j) measures their dissimilarity and determines the possibility as a cluster. Here, we adopt Euclidean distance. Some studies consider network constraint as an improvement in trajectory clustering, however, in this study, as the interest is in "area", Euclidean distance is more proper.

We take the single-linkage or nearest neighbor clustering criterion, which is one of the most widely used clustering criterias:

$$Dist(c_n, c_k) = Min(Dist(c_i, c_k) + Dist(c_j, c_k)) \tag{11}$$

where two clusters C_i and C_j are merged to generate a new higher level cluster C_n, and C_k is the remaining cluster.

Algorithm 1 Clustering Algorithm

```
Require:
  matrix D contains all distances d(i, j).
Ensure:
  L(k) is the level of the k^th clustering;
  the proximity between clusters (n) and (k) is denoted as
  d[(n), (k)].

  Level L(0) = 0;
  sequence number m = 0;
  while objects in more than one clusters do
    Find the least dissimilar pair of clusters in the current clustering;
    d[(n), (k)] = mind[(i), (j)];
    Cluster(m)=cluster(n) + cluster(k);
    Merge clusters (n) and (k) into a single cluster to form the next
    clustering m;
    L(m) = d[(n), (k)];
    m = m + 1;
    Update matrix D;
  end while
```

Hence, the distance between two clusters is computed as the distance between the two closest elements in the two clusters. As a result, clusters may be forced together due to single elements being close to each other, even though many of the elements in each cluster may be very distant to each other. This is called as "chaining phenomenon", and is usually treated as a drawback of single-linkage method. However, the bus information is constrained by linear road links which makes the most distinctive feature comparing with other spatially distributed data. Conventional clustering algorithms, even the emerging spatio-temporal clustering method [9] do not take linear distribution into consideration. Based on the notion that vehicles driving on the same road links usually share similar destination, trajectories along the same road link thus are more similar compared with those distributed on different road links even if with smaller spatial distance. Therefore, more weight is given to the trajectory points on the same road link, which exactly conforms to the chaining phenomenon.

Moreover, single-linkage clustering is deterministic, in the sense that the resulting clusters do not depend on the order in which elements having equal distances are chosen. This is not necessarily true of other linkage schemes. The clustering algorithm is described in Algorithm 1.

VI. Experimental Results

In this section, we first describe the experimental setup developed to test our model on the available data, then present and discuss the results.

A. Experimental Analysis of Passenger Demands

Our model produces an online forecast for the passenger demands in all bus stops at each P-minutes period. Such test is through an offline continuous simulation. The scripts used are developed using the R statistical software. The predefined functions used and the values set for the model's parameters are detailed along this section.

 ISBN: 978-1-61208-383-4

TABLE I. ACCURACY OF MODELS USING ALPHA(α)=0.4.

Model	Periods			
	5am to 9am	9am to 1pm	1pm to 5pm	5pm to 9pm
Poisson Mean	75.83%	71.69%	74.13%	73.58%
W. Poisson Mean	76.98%	74.27%	75.76%	75.18%
Arima	78.35%	73.79%	76.28%	75.99%
Ensemble	79.26%	76.59%	78.39%	77.96%

TABLE II. ACCURACY OF MODELS USING ALPHA(α)=0.5.

Model	Periods			
	5am to 9am	9am to 1pm	1pm to 5pm	5pm to 9pm
Poisson Mean	75.83%	71.69%	74.13%	73.58%
W. Poisson Mean	76.63%	73.58%	74.86%	74.39%
Arima	78.35%	73.79%	76.28%	75.99%
Ensemble	79.37%	76.62%	78.33%	78.06%

Here, the aggregation period is set to 30 minutes (i.e., a new forecast is produced each 30 minutes; P=30) and a radius of 100 meters is used ($W = 100$). These parameters are set according to the average waiting time in bus stop ($<$ 21 minutes).

The previously described dataset is divided into a training portion of 16 weeks, and a test portion of 6 weeks. Both training and test set are composed by one time series per bus stop and each value tested on the period t is merged to the training set to generate the forecast on the period $t+1$, i.e. the real number of passenger demands count for each bus at a given bus stop along 30 minutes are considered for the next period forecast, and so on.

The time-varying Poisson averaged models (both weighted and non-weighted) are updated every 24 hours. A sliding window of 4 hours (H=4) is considered in the ensemble. The accuracy of each model is measured using the metric, which is also used to weight each model in the ensemble - *sMAPE*. Distinct results for two distinct values (0.4 and 0.5) of the parameter *alpha* (α in the weighted average) are presented below.

A total of 86,411 bus services are tested. The accuracy measured for each model is presented in Table I and Table II. The results are firstly presented per shift and then globally. The values presented below are calculated through an average weight of the accuracy obtained in each one of the time series (i.e., the accuracy of the forecast on the passenger demand for buses on each one of the 416 bus stops). Each accuracy is weighted according to the number of services demanded on the corresponding bus stop along all the test periods.

Each model presents accuracy above the 76% in both tables. The W. Poisson Mean and the Ensemble are the only ones affected by the changes on the *alpha* (α) parameter. The sliding window ensemble is always the best model in every shift and period considered, with an accuracy superior to 78%: 67,400 of the 86,411 total services are correctly forecasted in both time and space using an aggregation of 30-minutes periods.

B. Experimental Analysis of Attractive Areas

To discover attractive areas, or so-called hot regions, we choose those areas that have over 60 pick-up and drop-off points. This is because each time span is 4 hours, i.e., 240 minutes. The 60 points can ensure there emerges at least one pick-up or drop-off point in the cluster less than every 4-minute in average, which is dense enough. As described above these vehicles usually run in one out of four 4-hour shifts: 5am-9am, 9am-1pm, 1pm-5pm and 5pm to 9pm. Each time span corresponds to a cluster result, as summarized in Table III and Table IV. Table III shows the cluster results on Monday, while Table IV shows the cluster results on Saturday. From the total number of pick-up and drop-off points and the number of clusters, we can see that on a typical workday, people tend to be more active during the time span of 5am-9am and 1pm-5pm, while in a typical weekend, people tend to be more active in the afternoon and at night.

TABLE III. CLUSTERING RESULTS ON MONDAY.

Time Span	Total points	No. of clusters (n>60)
5am-9am	10,682	12
9am-1pm	7,205	32
1pm-5pm	9,420	28
5pm-9pm	8,793	21

TABLE IV. CLUSTERING RESULTS ON SATURDAY.

Time Span	Total points	No. of clusters (n>60)
5am-9am	5,481	36
9am-1pm	7,915	28
1pm-5pm	9,506	15
5pm-9pm	10,472	21

Figures 2 and 3 give an overview of the distribution of studied bus pick-up and drop-off points on workdays and on weekends separately. From the distribution of the clusters in two figures, it can be further observed that the distribution of the attractive areas follow certain pattern while vary crossing workdays and weekends. For example, there are some areas which remain attractive, though the sizes of them vary with time. There are also areas whose LoA are time-dependent. Moreover, we can see that people's activities are more concentrated on work places and living places on workdays, while places of amusement and schools are more attractive on weekends.

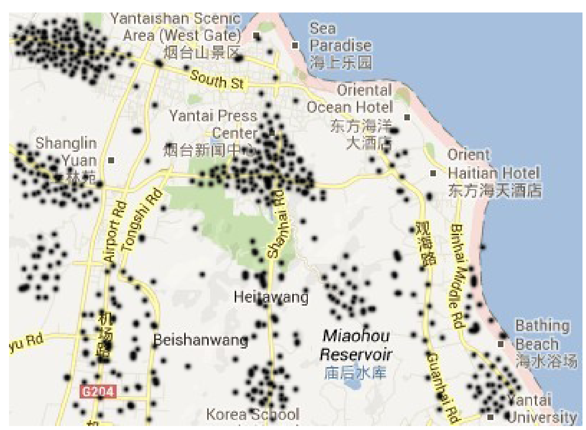


Figure 2. The cluster distribution on workday.

C. The Analysis of Movement Patterns of Passengers

The billing system of bus services records each passenger's behavior when a passenger gets on a bus and gets off a bus.

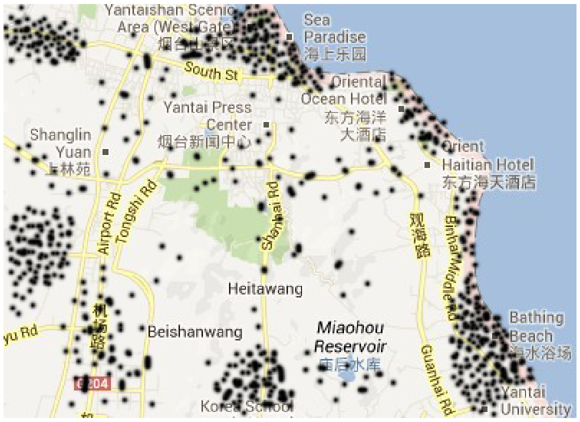


Figure 3. The cluster distribution on weekend.

Therefore, there is a certain correlation between the data of buses and the behavior of people in bus services. In order to study the correlation, we explore the data of buses to find the behaviors of people. The foregoing analysis shows that people who take a bus service may have the following characteristics:

- The service time is mainly about 15-30 minutes and only a minority is longer than an hour. The distribution of service time interval is approximated by a Power-law.
- In 77% cases, the number of passengers is over the available capacity of bus services. That means the buses are too crowded in most cases.
- Passengers' active time is often much concentrated. People tend to be more active during the time span of going to work and getting off work on workdays, while on weekends, people tend to be more active in the afternoon and at night.
- Attractive areas are time-dependent. People's activities are more concentrated on work places and living places on workdays, while places of amusement and schools are more attractive on weekends.

VII. Conclusion

In this paper, we present our work on mining spatial and temporal movement patterns of passengers on bus networks from three aspects: evaluating traveling time of passengers, predicting number of passengers to estimate passenger demand and congestion degree, and identifying attractive areas. The study is performed by transforming both GPS and event signals emitted by 1,326 buses in the City of Yantai, China into time series of interest. As a result, our method is able to identify attractive areas and predict passenger demand on buses at each one of the 416 bus stops at every 30-minutes.

Our method demonstrates a satisfactory performance, and predicts accurately more than 78% of the 86,411 demanded services, anticipating in real time the spatial distribution of the passenger demand. The approach is novel and can provide instructive insight to transport management, urban planning and location-based services. In particular, our work has major practical impact to help the management of bus companies to provide optimal services based on the knowledge of passenger movement patterns.

Acknowledgment

This research was partially supported by the Project of Shandong Province Higher Educational Science and Technology Program (No.J12LN05); the grants from the National Natural Science Foundation of China (No.61202111, 61273152, 61303017); the Project Development Plan of Science and Technology of Yantai City (No.2013ZH092); the Doctoral Foundation of Ludong University (No.LY2012023); the natural science foundation of Shandong province and special projects (Grant No.ZR2013FL009); the US National Science Foundation ACI 1443054 and IIS 1350885; Scientific Research Foundation for Returned Scholars of Ministry of Education of China (43th).

References

[1] M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi, ”Understanding Individual Human Mobility Patterns”. Nature 453, 2008, pp.779-782.

[2] M. Jordan, J. Kleinberg, and B. Scholkopf, ”Pattern Recognition and Machine Learning[C]”. Christopher M. Bishop. 2006.

[3] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, ”A Clusteringbased Approach for Discovering Interesting Places in Trajectories”. in SAC'08 Fortaleza, Ceara, Brazil, 2008, pp. 863-863.

[4] X. Li, C. Claramunt, C. Ray, and H. Lin, ”A Semantic-based Approach to the Representation of Network-constrained Trajectory Data”. In 12th International Symposium on Spatial Data Handling: Springer, 2006, pp. 451-464.

[5] L. O. Alvares, V. Bogorny, J. F. Macedo, B. Moelans, and S. Spaccapietra, ”Dynamic Modeling of Trajectory Patterns using Data Mining and Reverse Engineering”. In 26th International Conference on Conceptual Modeling - ER2007- Tutorials, Posters, Panels and Industrial Contributions. vol. 83, 2007, pp. 149-154.

[6] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. Macedo, F. Porto, and C. Vangenot, ”A Conceptual View on Trajectories”. Data and Knowledge Engineering vol. 65, 2007, pp. 126-146.

[7] F. Verhein and S. Chawla, ”Mining Spatio-temporal Association Rules, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases”. In 11th International Conference on Database Systems for Advanced Applications. vol. LNCS 3882, M. L. Lee, K. L. Tan, and V. Wuwongse, Eds.: Sringer-Verlag, 2006, pp. 187-201.

[8] C. Holt, ”Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages”. International Journal of Forecasting, 20(1), 2004, pp.5-10.

[9] D. Birant and A. Kut, ”ST-DBSCAN: An algorithm for clustering spatial-temporal data”. Data and Knowledge Enginnering. Vol 60, 2007, pp. 208-221.

[10] L. H. B.Williams, ”Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results”. Journal of Transportation Engineering, 129(6), 2003, pp.664-672.

[11] S. Makridakis, and M. Hibon, ”The M3-Competition: Results, Conclusions and Implications”. International Journal of Forecasting, 16(4), 2000, 451-476.

 ISBN: 978-1-61208-383-4