Real-Time Monitoring and Forecast of Active Population Density Using Mobile Phone Data

Qi Li(✉), Bin Xu, Yukun Ma, and Tonglee Chung

Department of Computer Science and Technology,
Tsinghua University, Beijing, China
zhongguoliqi@163.com, xubin@tsinghua.edu.cn, yukunma@aliyun.com,
tongleechung86@gmail.com

Abstract. Real-time monitoring and forecast of large scale active population density is of great significance as it can warn and prevent possible public safety accident caused by abnormal population aggregation. Active population is defined as the number of people with their mobile phone powered on. Recently, an unfortunate deadly stampede occurred in Shanghai on December 31th 2014 causing the death of 39 people. We hope that our research can help avoid similar unfortunate accident from happening. In this paper we propose a method for active population density real-time monitoring and forecasting based on data from mobile network operators. Our method is based solely on mobile network operators existing infrastructure and barely requires extra investment, and mobile devices play a very limited role in the process of population locating. Four series forecasting methods, namely Simple Exponential Smoothing (SES), Double exponential smoothing (DES), Triple exponential smoothing (TES) and Autoregressive integrated moving average (ARIMA) are used in our experiments. Our experimental results suggest that we can achieve good forecast result for 135 min in future.

Keywords: Real-time forecast · Population density · Public safety · Mobile phone data

1 Introduction

Active population density is vital in determining the possibility of accidents in public safety and there are many areas that has attracted a lot of attention, such as railway stations, coach stations, scenic spots and squares. However, abnormal population aggregation in areas that are least paid attention to poses a greater threat of causing safety problem in public. An effective way to monitor population at large scale in real-time and forecast the distribution of active population have not been proposed to our knowledge. In our research, we propose a method that makes use of mobile network operators existing facilities to achieve monitoring and real-time forecast of population density with high accuracy at large scale.

© Springer Science+Business Media Singapore 2016
W. Chen et al. (Eds.): BDTA 2015, CCIS 590, pp. 116–129, 2016.
DOI: 10.1007/978-981-10-0457-5_12

Mobile phones now have an extremely high penetration rate across the globe. The global mobile user penetration rate has reached 96 % in 2014 [1]. In developed countries, the number of mobile phone has surpassed the total population. In developing countries, it is as high as 90 % and continuing to rise. Because of the high penetration rate of mobile phone usage, the numerous types of data generated by mobile phone have become a promising source of data for researchers.

Network-based and handset-based localization technologies are two main categories of wireless location. Handset-based technology, which is based on sensors on the mobile phone such as GPS receiver, have a relatively higher location accuracy. However, high accuracy location collection sometimes violates privacy rights of mobile users, and users are not willing to share their location data unconditionally. Large scale participation of users in location recording is not applicable. In this paper, we propose a real-time active population density monitor and forecast framework which is based on network-based data and technology. Cellular network services are provided by base stations and each base station covers an area with a fixed radius. To provide high quality service, network operator usually build more base stations in densely populated areas and less on lower populated areas. Moreover, network operator collect data from all base stations, which include name, id, latitude, longitude, communication technology type etc. When mobile users use their cell phones to make a call, send a message or access the Internet, a cell phone will choose the nearest and most suitable service provided by surrounding base station. When a user moves to a new area, the cell phone will re-choose an appropriate service automatically according to signal strength and network standard. Network operator then store the records, which includes user identification, base station identification, time stamp etc., of communication between cell phone and base station to database or other storage devices.

A kind of Received Signal Strength Indication (RSSI) data was used in our experiments, also known as measurement report (MR) data. MR data is collected by digital cell mobile communications network sent by users' equipment. MR data is generated by each mobile devices every 480 ms. Because of this high frequency in data, trajectory of each user has a strong continuity in terms of time and space. So the MR data has very high research value for researchers.

In this paper, we propose a method for real-time monitoring and forecasting of active population density. We use Thiessen polygon to divide the coverage of service provided by each base station and calculate the area of the service coverage. We find a way to display population density by combining Thiessen polygon and population data provided by base stations. Four series forecasting methods,namely Simple Exponential Smoothing (SES), Double exponential smoothing (DES), Triple exponential smoothing (TES) and Autoregressive integrated moving average (ARIMA) are used in our experiments. We find that simple methods like SES and DES have a performance of validity of over 0.8 when time span is less than an hour with threshold set to 0.4. We can also see the ARIMA is better at forecasting active population density with longer time span.

The rest of the paper is organized as follows. In Sect. 2, we describe related work on forecast of population density. In Sect. 3, we give an overview of the

experimental data. Section 4 describes the four methods we use in the experiments. Section 5 shows experimental results, and gives detail analysis of the result. Finally we conclude this paper in Sect. 6.

2 Related Work

Until now, research and information on the area of population distribution and population density remains scarce [2]. However, both of these are very important to government policy making, urban management [3], academic research [4–6] and other fields. The development of techniques, such as simple areal weighting, dasymetric model [7–10], improves accuracy [7,10–15] and reduces workload. Despite these improvement, previous methods are constrained by population data from censuses reports [7,8,10–15]. In fact, because of the mobility of people, the number of people within a region is impossible to remain unchanged, and this change may occur very fast, both regularly or non-regularly. The population density of the area also changes rapidly correspondingly. So we may acquire some data at a certain time, but such data are out-dated already that they cannot meet our need. Recently, technologies based on cellular network [16] which make use of Call Detail Record (CDR) collected by the network operators and location information of base stations, are used to estimate the population and can produce a higher precision. CDR are data records produced by user when making a call. The estimation of population based on CDR cannot reflect the actual population and it also cannot be used to forecast future population.

Our research is also based on cellular network, but the data we used in our experiments is MR data which is collected by mobile network operators from mobile phone every 480 ms. The active population can be reflected precisely without extra methods or simulations. The MR data is very suitable for active population forecast.

3 Data Description

In this research, two different datasets are used in our experiments. Measurement report (MR) data, which are a kind of RSSI data sent by user equipment, are collected by digital cell mobile communications network. MR data generated by a cellphone usually contains the signal strength data measured from surrounding base stations and MR data are sent by mobile equipment to the base station which provides the communication service. Network operator can gather all the MR data from all the base stations. MR data are generated every 480 ms, so the high frequency in data, trajectory of each user has a strong continuity in terms of time and space. Our dataset contains 90720 records collected from two base stations from February 2015 to March 2015. Each record has been preprocessed and only contains useful features that can be expressed as a tuple in the form $[phoneID, stationID, timestamp]$. The interval between each record in our dataset is 1 min and time span of each record is 480 ms.

Another dataset contains 9,136 records of locations of distinct base stations in Hainan Province. Each record can be expressed using a tuple in the form $[stationID, latitude, longitude]$.

In our method, it is unnecessary to get precise position information of each person and the population in service coverage of each station is our concern.

4 Real-Time Active Population Monitoring and Forecasting

The main research goal of this paper is population density forecast. Population density is generalized as:

$$\rho = \frac{N}{S} \tag{1}$$

where, N is the population within service coverage and S is the area of service coverage. So population statistics and area of service coverage calculation are two important parts of our work.

4.1 Area of Service Coverage for a Base Station

Because of the uneven distribution of population, more base stations are built in areas with higher population density than in that with lower population density. Thus, the distribution of service coverage of base stations is not a standard honeycomb shape. Figure 1 shows a demonstration of a standard honeycomb shaped cellular network. Theoretical base station locations does not help us calculate the actual base service coverage area. An alternative way to get service coverage areas is by Thiessen polygon [17,18]. Figure 2 shows how service coverage can be divided by Thiessen polygon where each polygon represents the service coverage area of a base station.

The essence of Thiessen polygon is simply polygon, thus the process of getting the area of a Thiessen polygon is as follows. First, we divide the Thiessen polygon into some triangles by connecting one vertex to all the other vertices of polygon. Second, we convert the latitude and longitude coordinates into geometric distance. We assume the vertices coordinates are represented as $(latitude1, longtitude1)$ and $(latitude2, longtitude2)$. The geometric distance d of two coordinates is calculated using the formula:

$$d = 2 \times R \times \sqrt{\sin^2(\frac{a}{2}) + \cos(radLat1) \times \cos(radLat2) \times \sin^2 \frac{b}{2}} \tag{2}$$

where, R is earth radius, and

$$radLat1 = \frac{latitude1 \times \pi}{180.0} \tag{3}$$

$$radLat2 = \frac{latitude2 \times \pi}{180.0} \tag{4}$$

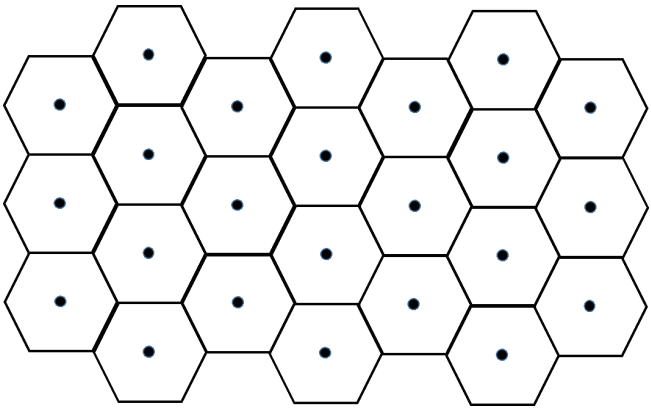

Fig. 1. Standard Cellular Network.

Fig. 2. Actual service coverage divided by Thiessen polygon in cellular network.

$$radLng1 = \frac{longtitude1 \times \pi}{180.0} \tag{5}$$

$$radLng2 = \frac{longtitude2 \times \pi}{180.0} \tag{6}$$

$$a = radLat1 - radLat2 \tag{7}$$

$$b = radLng1 - radLng2 \tag{8}$$

Based on the assumption that the earth is a standard sphere, we can obtain the actual geometric distance between two geometric coordinates using the above conversion. Finally, after calculating all the sides of the triangle, the Heron's

formula is applied to get the area of every divided triangles. Where the Herons formula is as follows:

$$s = \sqrt{p \times (p-a) \times (p-b) \times (p-c)} \tag{9}$$

$$p = \frac{1}{2}(a+b+c) \tag{10}$$

where a, b and c represents the three sides of one triangles respectively. Ultimately, the area of Thiessen polygon can be obtained by a summation of all the divided triangles.

$$area = s_1 + s_2 + \cdots + s_n \tag{11}$$

where $s_i (i = 1, 2, \ldots, n)$ represents the area of the ith triangles in Thiessen polygon.

4.2 Active Users Within Service Coverage of a Base Station

Measurement report (MR) is sent to a base station from a cell phone every 480 ms, so a base station can receive measurement reports from all cell phones within its service area in 480 ms. If mobile phones are turned on in the service area, they must send measurement reports back to the base station following service provider's requirement. So the number of mobile users calculated by parsing MR data is equal to the real number of mobile users whose mobile phone is powered on.

The Number of Active Mobile Users. We define mobile phone users with their phone turned on as active mobile users. People can either be in resting state or non-resting state in their daily lives. People with non-resting state have a larger range of activities so they are the main objectives of our research. The power on and power off state of a mobile phone have a strong connection with its user's state. When a mobile user is ready to have a rest or take a break and do not want be disturbed, such as sleeping at night, he or she will turn off the mobile phone. In this situation, the number of mobile users collected should be smaller than the actual number and this is normal, also this part of mobile users are active mobile users. It is usually that at a non-resting time or in a non-resting venue, mobile phones are usually turned on, so the number of active mobile users are almost the same as the actual number of mobile users. Therefore the number of active mobile users is our main research point.

Two Data Sets of Active Population. We obtain two datasets as our experimental data. Figure 3 shows our first dataset *MRDS1*, which is the coverage of a base station which is most probably a residential areas. February 18th 2015 is the eve of Chinese New Year. All family members will gather to celebrate the arrival of the New Year at home. At 20:00, the number of people reaches its peak. And at the beginning of February 19, there are still many people in active state. This makes sense if we understand the activity schedule of a typical Chinese family

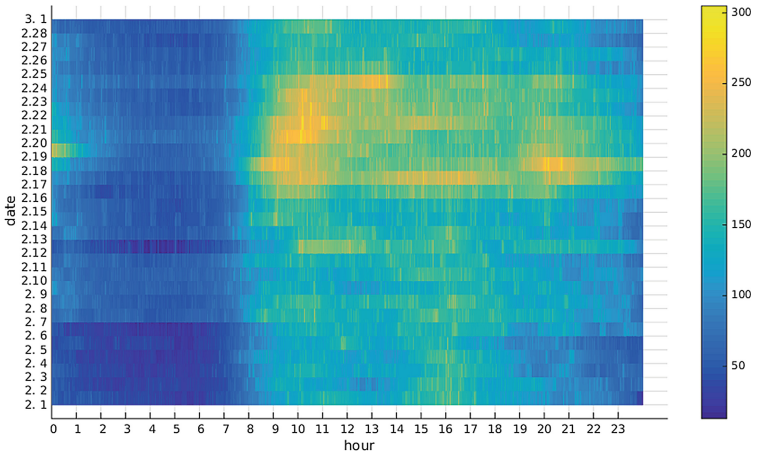

Fig. 3. *MRDS1* is a dataset of the population over a period of 29 *days* × 24 *hours*. The y axis refers to the date and x axis refers to the time in hours. A darker data point refers to a lower population number and vice versa.

on Chinese New Year. A strong activity presence is displayed between 9:00 and 22:00 from February 19 to February 25, this period is the official Chinese New Year statutory holiday.

Figure 4 shows our second dataset *MRDS2*, that is collected from a base station near shopping malls. On February 18th, activity presence reached a minimum, and in the following 7 days Chinese New Year statutory holidays, activity existence is at a very low state. In the other periods, it has a high degree of activity presence. Overall, the two graphs shows two obvious opposite trends.

The two figures accurately describe the activity pattern of Chinese people during the Chinese New Year holiday in two different areas.

4.3 Real-Time Monitoring and Forecast of Active Population Density

The active population data can be expressed in time series as shown in Fig. 5. $t_{current}$ is the current time and t_{latest} is the time of the latest data that can be observed while data before t_{latest} can also be observed. The situation after time $t_{current}$ is what we want to estimate and are interested in, which can be forecast according to history data. As we know, collecting and processing of data requires time, so the data that we can observed is history data and the delay time $T_{delay} = t_{current} - t_{latest}$. If we want to forecast the population density at time t, the time span is $T_{span} = t - t_{latest}$. Then we define that the actual value at time t as V_t.

Based on the above, for a base station, we can obtain the area of service coverage and the active population. The active population density can be calculated accurate.The real-time monitoring of active population density can

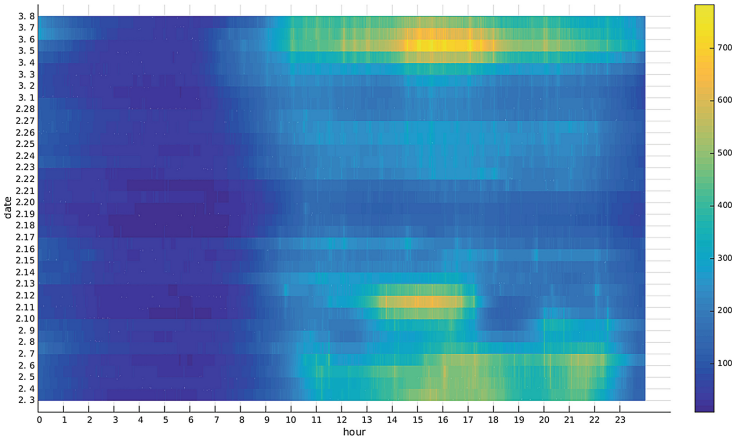

Fig. 4. *MRDS2* is a dataset of the population over a period of $35\,days \times 24\,h$. The y axis refers to the date and x axis refers to the time in hours. A darker data point refers to a lower population number and vice versa.

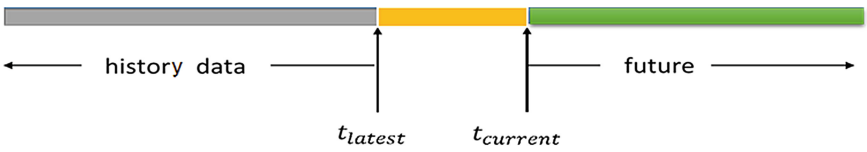

Fig. 5. Description of a time series in our experiment.

be done easily at a large scale. Obviously, the latest data we can observed at time t_{latest}.

The definition of real-time population density forecast is as follows:

If a method $F(t_k)$ uses the data before t_{latest} to forecast data at the future time $t_k (t_k \geq t_{current})$. The error of the forecast value and actual value $err_{T_{span},k}$ is define as the actual absolute value of the difference between the forecast value $F(t_k)$ and the actual value V_{t_k} divide by the actual value V_{t_k}. We use an indicator function 1_k that output 1 if $err_{(T_{span},k)}$ is less than or equal to a threshold ϵ and 0 if greater than threshold ϵ. Finally, we calculate a validity rate p of our forecast by taking a mean of the sum of all indicator function $1(k)$, $k = 1, \ldots, n$.

A formal definition can be expressed in following formulas:

$$err_{(T_{span},k)} = \frac{|F(t_k) - V_{t_k}|}{V_{t_k}} \tag{12}$$

$$1(k) = \begin{cases} 1, err_{(T_{span},k)} \leq \epsilon \\ 0, err_{(T_{span},k)} > \epsilon \end{cases} \tag{13}$$

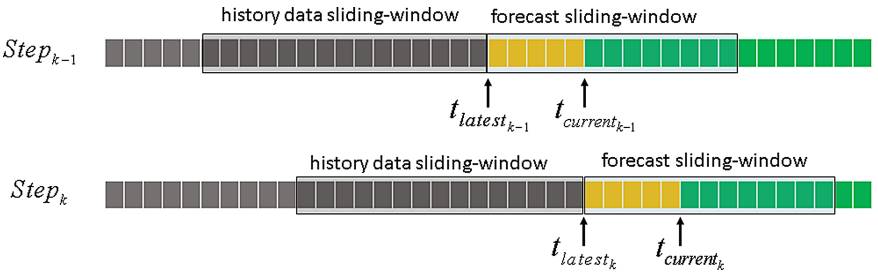

Fig. 6. Demostration of sliding window on a time series data.

$$p = \frac{\sum_{k=1}^{n} 1(k)}{n} \tag{14}$$

conditioned by,

$$t_{current} < t_1 < t_2 < \cdots < t_n \tag{15}$$

$$t_{current_k} \geq t_{latest_k} \tag{16}$$

where, $T_{span} = t_k - t_{current}$ and ϵ is the threshold.

Obviously, there are some relations between p and T_{span}, and we hope that p is as large as possible.

Figure 6 shows more detail. As we can see, we use a fixed sliding window is used in our experiment. The width of history data sliding-window W_h is fixed and time of each step is set to T_{delay}. Multiple time points after $t_{current}$ will be detected according to definition of real-time forecasting.

4.4 Forecast Methods

Four series forecasting methods,namely Simple Exponential Smoothing (SES), Double exponential smoothing (DES), Triple exponential smoothing (TES) and Autoregressive integrated moving average (ARIMA) are used in our experiments. Next, we will give a brief introduction to each these methods.

Exponential Smoothing. Exponential smoothing is usually applied to smooth data, which is used by many window functions in signal processing, acting as low-pass filters to remove high frequency noise.

1. Simple Exponential Smoothing (SES) [19,20]. The raw data sequence is represented by $\{x_t\}$ beginning at time $t = 0$, and the result generated by the algorithm is written as $\{s_t\}$. When the sequence begins at time $t = 0$,

$$s_0 = x_0 \tag{17}$$

$$s_t = \alpha x_t + (1 - \alpha) s_{t-1}, t > 0 \tag{18}$$

where α is the smoothing factor, and $0 < \alpha < 1$.

2. Double Exponential Smoothing (DES) [21]. The raw data sequence is represented by $\{x_t\}$, beginning at time $t = 0$. $\{s_t\}$ is the smoothed value for time t , and $\{b_t\}$ is the best estimate of the trend at time t. F_{t+m} is an estimate of the value of x at time $t + m(m > 0)$ based on the raw data up to time t. Double exponential smoothing is defined to be

$$s_1 = x_1 \tag{19}$$

$$b_1 = x_1 - x_0 \tag{20}$$

And for $t > 1$ by

$$s_t = \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1}) \tag{21}$$

$$b_t = \beta((s_t - s_{t-1}) + (1 - \beta)b_{t-1}) \tag{22}$$

where α is the data smoothing factor, $0 < \alpha < 1$, and β is the trend smoothing factor, $0 < \beta < 1$.

To forecast beyond x_t

$$F_{t+m} = s_t + mb_t \tag{23}$$

the initial value b_0 can be set according to needs. F_0 is undefined and $F_1 = s_0 + b_0$.

3. Triple Exponential Smoothing (TES) [22]. A sequence of observations is $\{x_t\}$, beginning at time $t = 0$ with a cycle of seasonal change of length L. $\{s_t\}$ is the smoothed value of the constant part for time t. $\{b_t\}$ is the sequence of best estimates of the linear trend that are superimposed on the seasonal changes. $\{c_t\}$ is the sequence of seasonal correction factors. $\{c_t\}$ is the expected proportion of the forecast trend at any time t mod L in the cycle that the observations take on. The output of the algorithm is F_{t+m}, an estimate of the value of x at time $t + m$, $m > 0$ based on the raw data up to time t.

$$s_0 = x_0 \tag{24}$$

$$s_t = \alpha \frac{x_t}{c_{t-L}} + (1 - \alpha)(s_{t-1} + b_{t-1}) \tag{25}$$

$$b_t = \beta(s_t - s_{t-1} + (a - \beta)b_{t-1}) \tag{26}$$

$$b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1} \tag{27}$$

$$c_t = \gamma \frac{x_t}{s_t} + (1 - \gamma)c_{t-L} \tag{28}$$

where α is the data smoothing factor, $0 < \alpha < 1$, β is the trend smoothing factor,$0 < \beta < 1$, and γ is the seasonal change smoothing factor, $0 < \gamma < 1$.

Autoregressive Integrated Moving Average (ARIMA) [23]. ARIMA is used to time series data either to better understand the data or to forecast future points in the series. It can be defined as following,

$$Y_t = (1 - L)^d X_t \tag{29}$$

$$(1 - \sum_{i=1}^{p} \phi_i L^i) Y_t = (1 + \sum_{i=1}^{p} \theta_i L^i) \varepsilon_t \tag{30}$$

where X_t is a raw data sequence, t is an integer index, L is the lag operator, the θ_i are the parameters of the moving average part, The error terms ε_t are independent, identically distributed variables sampled from a normal distribution with zero mean. parameters p, d, and q are non-negative integers, p is the order of the Autoregressive model, d is the degree of differencing, and q is the order of the Moving-average model. ARIMA models are generally denoted $ARIMA(p, d, q)$.

5 Experiments

In this section, we describe our experiment in detail which includes our environment setting, our evaluation setting and our findings and results.

5.1 Environment Setup

We use two datasets for our experiments. One dataset consist of a time span of 34 days and the the other 29 days. The actual data collection frequency in application is one data per minutes. Altogether, we gather 90720 data points for our experiment. We set our history data sliding window W_h to 3000. The time of each step is set to T_{delay} to 15. The time for each step is set according to the data collection time cycle of the actual situation.

We test four different methods for forecasting, namely, simple exponential smoothing (SES), double exponential smoothing (DES), triple exponential smoothing (TES) and autoregressive integrated moving average (ARIMA).

5.2 Evaluation Method

We evaluate the methods using the error function that we defined in Sect. 4.3. We set four different threshold values $\epsilon = 0.05, 0.1, 0.2, 0.4$. Ten different forecast time is used, we choose $t_{span} = 15, 45, 75, 105, 135, 165, 195, 225, 255, 285$ respectively. We present our findings in the next section.

5.3 Results

Figure 7 shows there results for our first dataset MRDS1. We can that forecasting result for a time span of 15 min is satisfying. We can see that when we set ϵ to 0.4, validity rate for SES and DES is over 0.9. Even if we set ϵ to 0.05, we can see a validity rate of almost 0.5 for SES and DES. That means if we set the threshold to a strict condition, we can still get good results with some of the forecasting methods. The same results can be observed when time span is set 45 min. But we can see that the longer the time span, the harder it is to get good results. With time span over 135 min, the results are below expectation. We can conclude that our methods have a fair performance within 1 h. In other words,

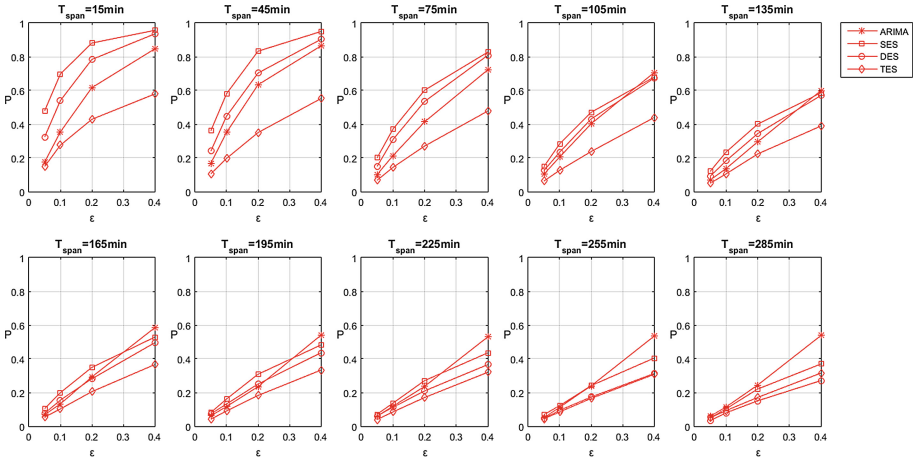

Fig. 7. Experimental results for MRDS1.

our forecasting methods can satisfy our requirement for real-time population density forecasting.

Figure 8 shows our result for the second dataset MRDS2. From our observation, the results are better than that of the MRDS1. Even with time span of 105 min, we can still obtain a satisfying result. But if we compare the two results, we find a similar trend, that is the longer the time span, the least accurate our forecast results. Moreover, we find that the simple methods like SES and DES have better performance when time span is small. Some other observations we find is that the more complex method ARIMA has a more stable performance

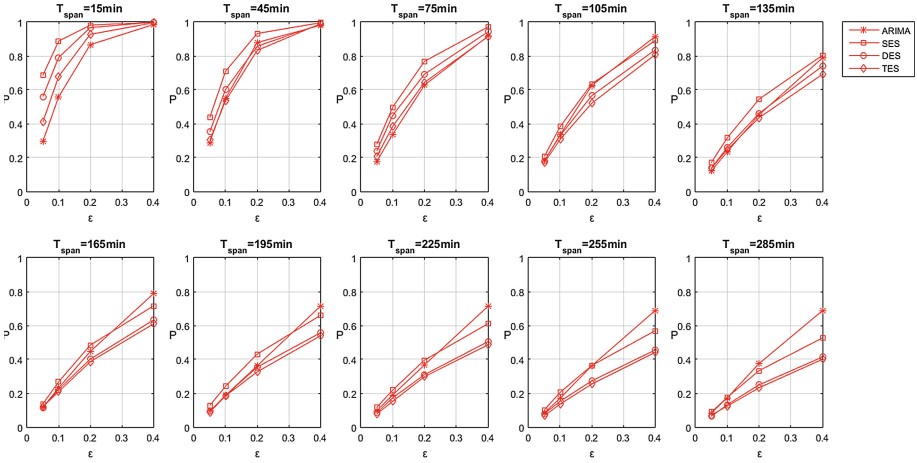

Fig. 8. Experimental results for MRDS2.

over time span, even when forecasting long time span, we can acquire good results to a certain degree.

6 Conclusion

In this paper, we propose a method to effectively monitor real-time active population density by using Thiessen polygon to divide mobile network based on coordinate data calculation and at the same time gather active population of each base station. Based on the above results, we also use four different methods to forecast active population density. From the results, we find that simple methods like SES and DES have a performance of validity of over 0.8 when time span is less than an hour with threshold set to 0.4. We can also see that the ARIMA has a stable performance at forecasting active population density with longer time span. The experimental results show that it is feasible to use the method of real-time monitoring and forecasting active population density in our research to avoid some possible public safety accidents.

Our work can be a foundation for more future work that uses complex models to better forecast active population density and for developing high level applications.

Acknowledgement. This work is supported by China National Science Foundation under grant No. 61170212, China National High-Tech Project (863) under grant No. 2013AA01A607, Ministry of Education-China Mobile Research Fund under grant No. MCM20130381, and Tsinghua University Initiative Scientific Research Program (No. 20131089190). Beijing Key Lab of Networked Multimedia also supports our research work.

References

1. World telecommunication development conference (wtdc-14): Final report. Technical report, International Telecommunication Union, (ITU, Dubai, United Arab Emirates) (2014)
2. Linard, C., Tatem, A.: Population mapping of poor countries. Nature **474**, 36 (2011)
3. Bongaarts, J., Sinding, S.: Population policy in transition in the developing world. Science **333**, 574–576 (2011)
4. Linard, C., Tatem, A.J.: Large-scale spatial population databases in infectious disease research. Int. J. Health Geogr. **11**(7) (2012)
5. O'Neill, B.C., Dalton, M., Fuchs, R., Jiang, L., Pachauri, S., Zigova, K.: Global demographic trends and future carbon emissions. Proc. Nat. Acad. Sci. **107**(41), 17521–17526 (2010)
6. OLoughlin, J., Witmer, F.D.W., Linke, A.M., Laing, A., Gettelman, A., Dudhia, J.: Climate variability and conflict risk in East Africa, 1990–2009. Proc. Nat. Acad. Sci. **109**(45), 18344–18349 (2012)
7. Balk, D., Yetman, G.: The global distribution of population: evaluating the gains in resolution refinement. New York: Center for International Earth Science Information Network (CIESIN), Columbia University (2004)

8. Deichmann, U., Balk, D., Yetman, G.: Transforming Population Data for Interdisciplinary Usages: From Census to Grid. Center for International Earth Science Information Network, Washington DC (2001)
9. Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J.: Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PloS one **10**(2), e0107042 (2015)
10. Gaughan, A.E., Stevens, F.R., Linard, C., Jia, P., Tatem, A.J.: High resolution population distribution maps for Southeast Asia in 2010 and 2015. PloS one **8**(2), e55882 (2013)
11. Tobler, W., Deichmann, U., Gottsegen, J., Maloy, K.: World population in a grid of spherical quadrilaterals. Int. J. Popul. Geogr. **3**(3), 203–225 (1997)
12. Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A.: Landscan: a global population database for estimating populations at risk. Photogram. Eng. Remote Sens. **66**(7), 849–857 (2000)
13. Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A.: Determining global population distribution: methods, applications and data. Adv. parasitol. **62**, 119–156 (2006)
14. Linard, C., Gilbert, M., Snow, R.W., Noor, A.M., Tatem, A.J.: Population distribution, settlement patterns and accessibility across Africa in 2010. PloS one **7**(2), e31743 (2012)
15. Azar, D., Engstrom, R., Graesser, J., Comenetz, J.: Generation of fine-scale population layers using multi-resolution satellite imagery and geospatial data. Remote Sens. Environ. **130**, 219–232 (2013)
16. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., Tatem, A.J.: Dynamic population mapping using mobile phone data. Proc. Nat. Acad. Sci. **111**(45), 15888–15893 (2014)
17. Aurenhammer, F.: Voronoi diagrams survey of a fundamental geometric data structure. ACM Comput. Surv. **23**(3), 345–405 (1991)
18. Okabe, A., Boots, B., Sugihara, K.: Spatial Tessellations. Concepts and Applications of Voronoi Diagrams. John Wiley & Sons, New York (1992)
19. Holt, C.C.: Forecasting seasonals and trends by exponentially weighted moving averages. Int. J. Forecast. **20**(1), 5–10 (2004)
20. Brown, R.G.: Smoothing, Forecasting and Prediction of Discrete Time Series. Courier Corporation, New York (2004)
21. Kalekar, P.S.: Time series forecasting using holt-winters exponential smoothing. Kanwal Rekhi Sch. Inf. Technol. **4329008**, 1–13 (2004)
22. Winters, P.R.: Forecasting sales by exponentially weighted moving averages. Manage. Sci. **6**, 324–342 (1960)
23. Makridakis, S., Hibon, M.: Arma models and the boxcjenkins methodology. J. Forecast. **16**(3), 147–163 (1997)