MODELLING DYNAMIC PATTERNS USING MOBILE DATA

Suhad Faisal Behadili[1], Cyrille Bertelle[1] and Loay E. George[2]

[1]Normandie Univ, LITIS, FR CNRS 3638, ISCN, ULH, Le Havre, France
`suhad.behadili@etu.univ-lehavre.fr`
`cyrille.bertelle@univ-lehavre.fr`
[2]Baghdad University, Computer Science Department, Baghdad, Iraq
`loayedwar57@scbaghdad.edu.iq`

ABSTRACT

Understanding, modeling and simulating human mobility among urban regions is very challengeable effort. It is very important in rescue situations for many kinds of events, either in the indoor events like evacuation of buildings or outdoor ones like public assemblies, community evacuation, in exigency situations there are several incidents could be happened, the overcrowding causes injuries and death cases, which are emerged during emergency situations, as well as it serves urban planning and smart cities. The aim of this study is to explore the characteristics of human mobility patterns, and model them mathematically depending on inter-event time and traveled distances (displacements) parameters by using CDRs (Call Detailed Records) during Armada festival in France. However, the results of the numerical simulation endorse the other studies findings in that the most of real systems patterns are almost follows an exponential distribution. In the future the mobility patterns could be classified according (work or off) days, and the radius of gyration could be considered as effective parameter in modelling human mobility.

KEYWORDS

Modelling, Armada, Probability Distribution, Inter-event time, Displacements

1. INTRODUCTION

The purpose of simulation analysis is to acquire and analyze the results in well conceptual vision, in order to give high indications for decision makers, pivoting on two events kinds, the discrete events and frequent (continuous) events. In order to explore the mobility characteristics, eventual real effects conditions, and actions of a specified system, they should be modeled mathematically [1, 2, 3, 4, 5]. However, the most important and difficult issue in modeling and simulating any system is the determination of the probability distribution and parameters to model the uncertainty of the system input variables. Many researches used CDRs to study the collective and individual human mobility (Gonzalez et al., 2008), the segmentation of urban spaces (Reades et al. 2009), understand of social events (Calabrese et al., 2010). (Phithakkitnukoon et al., 2010) suggested the activity-aware map, using the user mobile to uncover the dynamic of inhabitants, for urban planning, and transportation purposes. (Ratti et al., 2010) deal with large telecommunication database for Great Britain in order to explore human interactions, and emphasized on the highly interactions correlation with administrative regions [10].

2. DATA SET

The case study data is Call detail Records of mobile phone, composed of 51,958,652 CDRs, represent entry records of 615,712 subscribers, for the period starting from 4 (Friday)-15 (Tuesday) of July in 2008. However, it contains individuals occurrence in discrete (irrelevant) mode only, means that any mobile individual activity is recorded at (start/end) time, but there is a lack of information, which is supposed to indicate the individual's occurrence during inactive case (mobility without any mobile phone activity). There is no data meanwhile the mobile phone is idle, i.e. inactive or doesn't make any communication activities neither calls nor SMS activities. As well as, the available spatial data is only of the towers (X,Y) coordinates, hence it would be considered to estimate individual's transitions from position to another (from tower to tower), hence the positions would be determined approximately with regarding to the tower coverage (signal strength). With regarding to the non-deterministic & discrete nature of this data, therefore the collective behavior would be the effective approach to be analyzed and simulated. Since, each individual could be disappeared for a while from the DB, which makes individual tracing is unworthy, without significant indications on the people behavior in the city [3].

3. DESCRIBING INDIVIDUALS' ACTIVITIES

The communications activities have heterogeneous nature, since the individuals are varied in their usage of mobile phone, which are ranged between (rarely-frequently) usages during specific period. The individuals are grouped according to their total activities. However, probability of waiting time (inter-event time) ΔT of each two consecutive activities has been computed for each individual, so the individuals would be grouped with regarding to their activities. Computing the probability function is to get the system universal behavior (pattern), according to consecutive inter-event times, where most life systems are modeled by exponential law [1, 6, 7, 5, 8]. So, the distribution of the average inert-event time ΔT_a is estimated by exponential distribution law as in equation (1), its histogram in figure 1. The demonstration in figure 2, reveals that the longer waiting times are characterizes the individuals of fewer activities.

$$P(T) = (\Delta T)exp^{-\Delta T} \qquad\qquad (1)$$
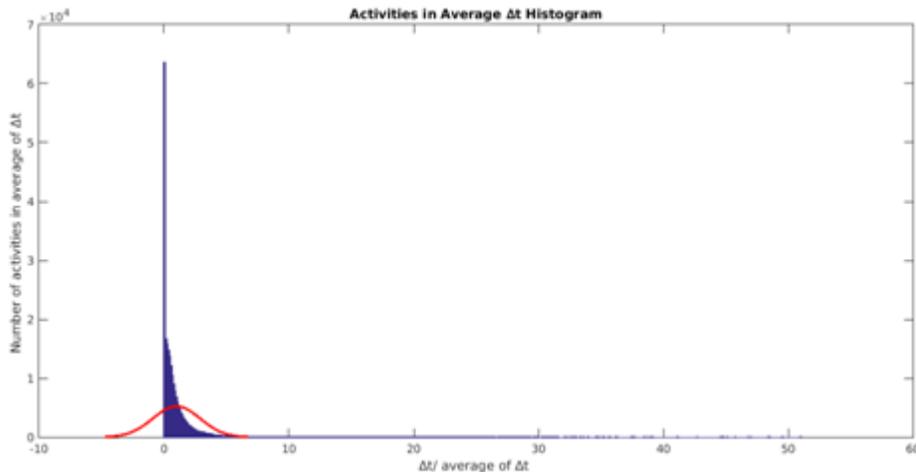
Figure 1: Histogram of Activities in Average of Inter-Event Time

4. INTER-EVENT TIME OBSERVATIONS

The modeling and simulation process starts by computing probability model, which is modeled the data and simulate the responses of the model, using one of the well-known functions, the probability density function PDF [3, 4, 9]. The main parameter in this study is the mean of inter-event time, where the means of samples of the distribution are drawn as exponential distribution, then compute the mean of all means. In order to understand the behavior of sample means from the exponential distribution, to get the universal system pattern (general population law). The computations are done as follows:

1. Manipulate all 12 days data, each day independently due to daily regular patterns, for all users (individuals) in spatio-temporal manner. Then eliminate the individuals of only one occurrence in the CDRs, since they didn't have significant indication on mobility.

2. Sorting the data by time, in order to have the real sequence of positions transitions of individuals' trajectories. Classify the data according to individuals' activities (sampling), by computing inter-event time Δt (waiting time), where it is the time elapsed between each successive activities of each individual, which is ranged between 15-1440 minutes, this sampling is done according to logical intuition, since 15 minutes is the minimum time that can give mobility indication, and the 1440 minutes (24 hours) could be considered as the highest elapsed time to travel inside the observed region.

3. Compute the inter-event time of all individuals ΔT, then compute ΔT_a (average inter- event time) of all individuals, then classify (min, max) samples according to activities score (activities densities), then compute $P(\Delta T)$. However, $\Delta T/\Delta T_a$ is the average inter-event time of the all individuals (whole population).

4. Compute exponential distribution probability for each day, then for the whole days as in figure 2 to identify the universal population pattern law, then to show the approximation of all 12 days curves with the curve of their average values as in figure 3.
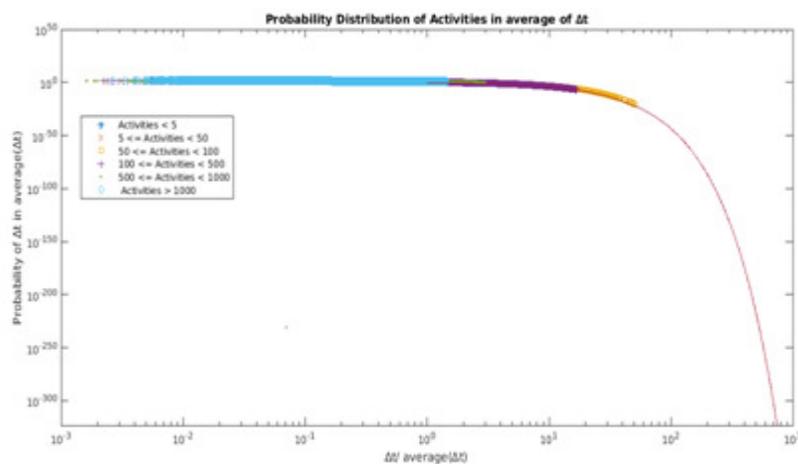
Figure 2: Waiting time distribution P(∆t) of mobile activities, where ∆t is spent time between each two successive activities, legend symbols are used to distinguish the individuals' groups according to their activities ratio, for whole population during whole period, activities in average ∆t.
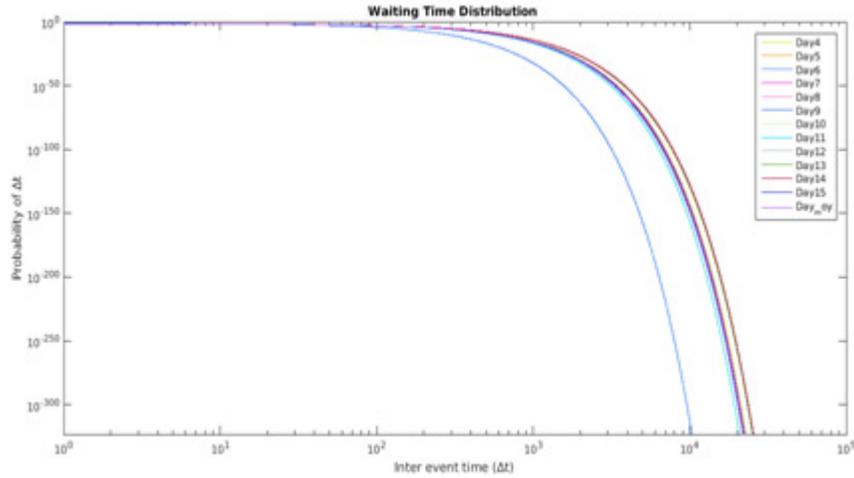
Figure 3: Waiting time for the 12 days curves with the curve of their average.

Exponential distribution (probability distribution) in equation (1) is capable of modeling the events happened randomly over time. In this case it describes the inter-event time, and the average of inter-event time of individuals' activities [1, 2, 3, 9]. Hence, the cutoff distribution is determined by the maximum observed inter-event time (significant parameters), which the individual can wait to make any mobile activities, where it is Δt=1431 min. The resulted law as in equation (1) is computed according to algorithm of complexity $O(n^3 + 2n^2 + n)$. The distribution shows that short inter-event times have higher probability than long ones, and the 12 days have similar patterns of activities. As well as, by following the same manner mentioned above, in order to compute the displacement statistics of all the individuals during the whole observed period, the displacement probability distribution is computed. However, the Δr is the traveled distance between each two successive activities during time ΔT_0 of the range 20-1440 minutes, and the P(Δr) is the displacement distribution, the investigated distances would be limited by the maximum distance that may be traveled by individuals in the ΔT_0 (time intervals). Hence, the cutoff distribution is determined by the maximum observed distance, which individual can travel is Δr=7.229515 Km along the day hours, since the maximum time slice couldn't exceed 24 hours with regards to observed region. The displacements distribution is approximated by power low as in equation (2). The resulted distribution shows as in figure 4 that the displacements are clear in 10^4 m, whereas decreased after this value, the Δr is ranged between (0-1) Km.

$$P(\Delta r) = (\Delta r)exp^{-(\Delta r)} \qquad\qquad (2)$$

Well, the traveled distance (displacements) distributions of the 12 days curves with the curve of their average are demonstrated in figure 5, they have almost identical patterns. Hence, the obtained averages of the waiting time distribution are max ΔT = 1431 minute and min ΔT =0 minute, whereas the max Δr =7.229515e+04 meter and min Δr =0 meter.
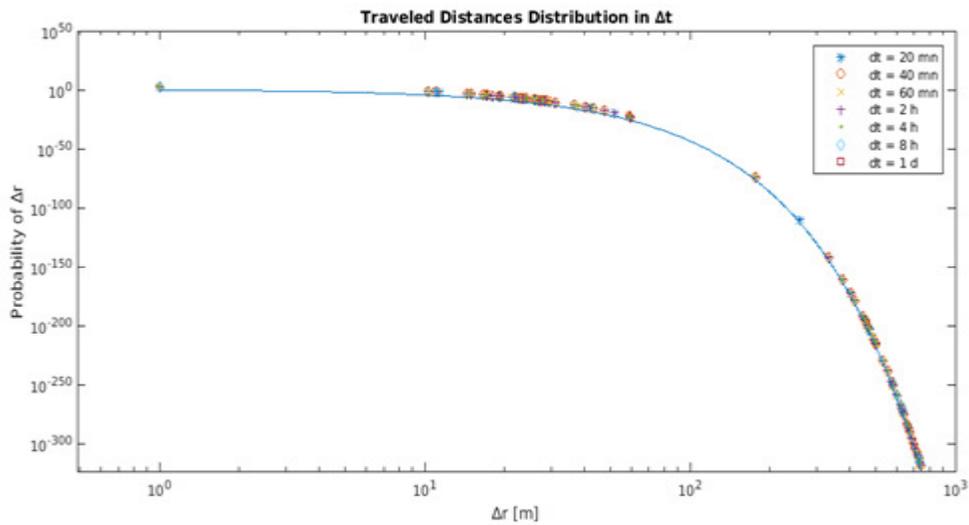
Figure 4: Distances probability distribution (displacements) P(Δr) for waiting times (inter-event times) Δt_0 for one day, cutoff distribution is determined by the maximum distance traveled by individuals for specific Δt_s.
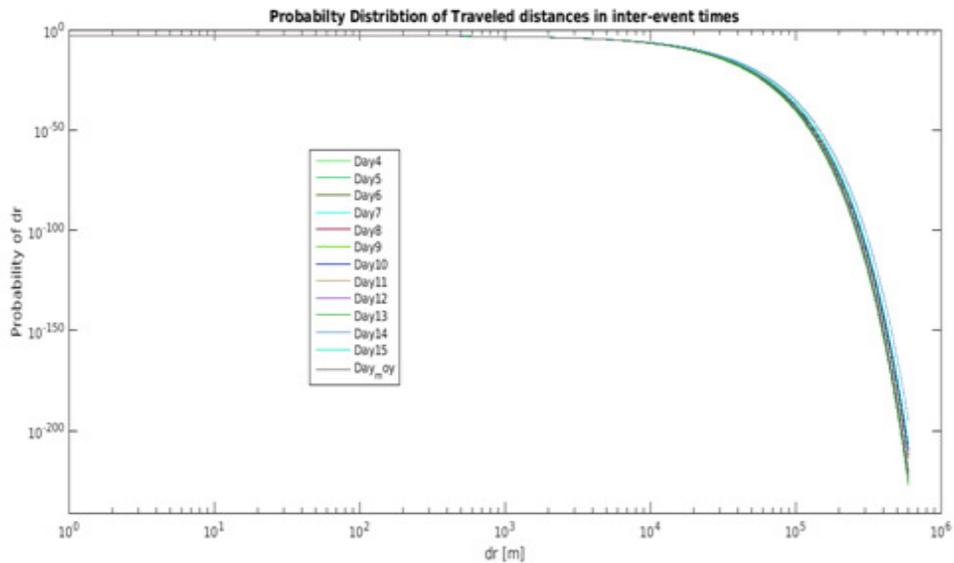
Figure 5: Displacements for the 12 days curves with the curve of their average.

5. CONCLUSIONS

The great attending recently towards dealing with catastrophic disasters, cities planning, diseases spreading, traffic forecasting, decision making for rescue people with disabilities in emergencies ...etc. [10], all these activities need working on huge data and a geographic data, to describe human behavior and mobility patterns, for giving the simulation model more reality, since almost simulation platforms are deal with only a grid to be the background of the model, so the resulted model will be far away from reality.

The results of this investigation show that the inter-event time Δt between each two successive activities has bursty pattern, since there is long period without activities, this gives indication about the population's heterogeneity. And human trajectories follows power-law distribution in step size Δr, and they are modeled using displacements and waiting time distributions, as well as the mobility traveling patterns are show that there are many short distances in contrast little long distances. The mobile phone data is used to reconstruct the population trajectories. However, the patterns of all days are very similar and they have approximately identical curves either in spatial or temporal features. As next forward step with this kind of studies the mobility patterns would be classified with regarding to the (work or off) days for more understandings to the life patterns. Whereas, the radius of gyration would be considered as significant modelling parameter to give the model more reality with focusing on patterns regularity.

REFERENCES

[1]   CS302 Lecture notes, (2015) "Simulating Random Event",
      http://web.eecs.utk.edu/~leparker/Courses/CS302-Fall06/Notes/PQueues/random_num_gen.html

[2]   "Simulation: An Introduction", (2015) Stat. Chapter 10,
      http://www.stat.nus.edu.sg/~zhangjt/teaching/ST2137/lecture/lec%2010.pdf

[3]   Sougata Biswas, (2014) "Simulation of Exponential Distribution using R",https://rstudio-pubs-static.s3.amazonaws.com/26693_e1151035722942b2813c0063c6b220ae.html

[4]   Vincent Zoonekynd, (2015) "Probability Distributions", http://zoonek2.free.fr/UNIX/48_R/07.html

[5]   Solver, (2015) "Simulation Tutorial Introduction", http://www.solver.com/simulation-tutorial

[6]   Siqi Shen, Niels Brouwers & Alexandru Iosup, (2011) "Human Mobility in Virtual and Real Worlds: Characterization, Modeling, and Implications", report number PDS-2011-007, ISSN 1387-2109.

[7]   The Pennsylvania State University, (2015) "Probability Theory and Mathematical Statistics", STAT 414/415, https://onlinecourses.science.psu.edu/stat414/node/97.

[8]   Ilya Narsky, Frank C. Porter, (2013) "Statistical Analysis Techniques in Particle Physics, Fits,Density Estimation and Supervised Learning", Wiley-VCH, 1 edition.

[9]   Marta C. Gonzlez, Cesar A. Hidalgo & Albert-Laszlo Barabasi, (2009) "Understanding individual Human Mobility Patterns", nature 453.

[10] Gunther SAGL, Bernd RESCH, Bartosz HAWELKA, Euro BEINAT, (2012) "From Social Sensor Data to Collective Human Behaviour Patterns-Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments", GI_Forum 2012: Geovizualisation, Society and Learning. Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-521-8.