Finding Probability Distributions of Human Speeds

Ha Yoon Song
Department of Computer Engineering,
Hongik University, Seoul, Korea
Email: hayoon@hongik.ac.kr

Jun Seok Lee
Department of Computer Engineering,
Hongik University, Seoul, Korea
Email: rolunoa@gmail.com

***Abstract*—For various research and industry purposes, knowledge of the characteristics of human mobility is required. In this paper, we will estimate the speed distribution of everyday human mobility using the unit of 0.01 m/s. From various smart mobile devices, a huge number of positioning data were collected, from which human mobile speed values are calculated. In a range of speed up to 108 Km/hour and 180Km/hour, we fit the speed data into probability distribution functions in order to establish a base for human mobility research, to which we believe this paper can make a significant contribution.**

Keywords*–*Speed distributions; Human speed; Human Mobility; Distribution Fitting.

I. INTRODUCTION

Monitoring human mobility is necessary for various purposes, including mobile computing and transportation engineering. Such work often requires natural science knowledge to ascertain the probability distributions of molecular mobility, wind velocity, and so on. The major purpose of this type of research is to model irregular motions into models with predictability at a certain level. Using the probability distribution of human speed, it is possible to predict human velocity, to detect abnormal motion, or to detect positioning system errors. If a monitoring system detects abrupt human mobility, this may imply an emergency or surveillance situation. There are a great number of methods for detecting abrupt human mobility. In addition, it is possible for an end-user to carry a mobile device that can report his or her position by use of embedded positioning functionalities and positioning systems. Such devices include dedicated GPS receivers and commercial smartphones, the latter using various combinations of positioning systems [1]. It is also possible to calculate distance between two points from positioning data, generally using the Haversine method [2]. From the distance and time, speed values may also be delivered from two consecutive positioning data.

The speed values impose distance information as well as the length of time interval between two positions, i.e., they are normalized values. In order to detect abrupt mobility, the speed value is key for calculating an abrupt change of position for a given time interval. However, there is at present inadequate knowledge of the probability distribution of speeds as related to everyday human life. Specifically, the question is, how can we scientifically define abrupt human mobility?

In this paper, we are going to analyze the probability distribution of speeds and will present several outstanding well-fits designed for practical use across a reasonable speed range found in everyday human life. Once we have found this distribution of speed, which forms the basis of human mobility, it will also be possible to calculate abrupt human mobility. Our aim is to provide the well-fit speed distribution of everyday human life.

In Section II, we will discuss the results of past research and will indicate the positioning data collection procedure used herein. Section III will show several important distributions that result from our fitting. We will conclude this paper in Section IV.

II. STATISTICAL BACKGROUND

There are about three previous research to be discussed as prerequisites for this research. The first one is previous research on the probability distribution for human mobile speed, the second one is positioning system technologies and positioning data collection, and the final one is data fitting between real world data and probability distribution.

A. Previous Research

There is very little research regarding human mobile speed and human mobile distance. In the very first research, cellular network-based location system used to collect mobile phone user's location data and analyzed mobile distance [3]. Cellular network-based location system can identify the location of mobile phone user based on cellular station and consecutive mobile user location data used to identify distance. It is found that the probability distribution function of human mobile distance is truncated power-law distribution. In addition Access Point (AP) of Wireless LAN (WLAN) can be used to analyze human mobility [4]. Pre-identified location of AP can be used and then the MAC address of each device in combination with the time analyzed to identify the mobile distance of each device. The result of this research is that human mobile distance follows log-normal distribution. Another research utilized GPS data of taxi [5]. This research reveals human mobile speed follows exponential distribution.

In our previous research [6], we tried to figure out proper probability distributions of human speeds in various categories. We tried to calculate the proper probability distributions of human speeds across various categories. The units of speeds were 0.1 m/s, 0.5 m/s and 1.0 m/s. However, this is somewhat inaccurate considering human micro-mobility. Once we have bigger unit then 1.0 m/s, it is hard to fit for continuous probability distributions. Unlike previous methodologies, we now collect more positioning data, using the more precise unit of 0.01 m/s. Our analysis is also carried out using the Kolmogorov-Smirnov test [7]. The fit between raw data and each candidate probability distribution can be found by K-S test.

ISBN: 978-1-61208-356-8

B. Data Collection

Positioning data sets were collected for this research. The longest collection period was from March 2013 to Feb 2014. Several individuals carried their mobile devices whilst in a resting or moving state outside of their own home. The area of collection was mostly the metropolitan area of Seoul, Korea, however it included other part of Korea, as well as countries such as the USA, Canada, France, Italy, the Netherlands, Austria. A total of 2,218,020 speed values were collected.

The devices used for collection were as follows: iPhone 3Gs, iPhone 4 [8], iPhone4S [9], Galaxy S3 [10], Galaxy Note2 [11], Garmin Edge 800 [12], Garmin EDGE 810 [13], Garmin 62s [14]. For dedicated devices, such as Garmin, no app is requrired; however, apps for collecting positioning data from smartphones are required. Such smartphones use hybrid positioning system rather than just GPS in order to figure out the position of devices. We thus developed positioning data collecting apps for iPhone and Android phones and some of the positioning data were also collected using commercial apps. Using a variety of collection methods was intentional since we needed to cope with every possible situation of positioning data collection to guarantee the generality of the data collection.

TABLE I. Number of data in unit of 0.01 m/s.

Speed Region (m/s)	Count
Total	2,218,020
0	355,832
0.01 - 50	1,853,045
0.01 - 30	1,836,608
2.78 - 50	860,672
2.78 - 30	844,235

As the actual speed data calculated had 12 digits below the point, we needed to prune out the data to achieve reasonable values. We chose units of 0.01 m/s, which correspond to 36 meter/hour. These values are likely to be precise enough for use as everyday human speed units. One of the clear phenomena contained in speed values is that the data set will contain a lot of zero speed values, i.e., the distribution is zero inflated. In such situations, even though speed value is not zero, but it is rather less than 0.01 m/s, such value will be treated as zero. Therefore, we decided to exclude zero values. And considering that a person can move anytime in a certain distance, for example 2.77 meter in a second, we also investigated the distribution so as to excluded speed values less than 2.77 m/s. In other words, we intentionally exclude speed values less than 2.78 m/s since we are considering only the meaningful mobile situation. We also needed to decide the upper speed bound. From various possible upper bounds, we chose the values of 30 m/s and 50 m/s, which correspond to 108 Km/hour and 180 Km/hour, respectively. We considered these values to represent reasonable limits for speed in everyday human life. We also conducted distribution fitting for the four categories of speed data. The frequency of each speed value was calculated and the Maximum Likelihood Estimation (MLE) [15] was applied in order to find the parameters of probability distribution, and then K-S statistics were obtained [7].

Table I is an overall summary of the number of data classified into four categories.

III. RESULTS

Table II - V present the results of fitting for each of the positioning data categories. Table II summarizes 20 distinguished distributions for the speed range of 0.01 m/s to 30 m/s. It contains the rank of probability distribution by K-S statistic, the name of the distributions, the K-S statistic values, and the parameters for the corresponding distributions. For example, lognormal distribution shows the best fit with statistic 0.03859 for the speed data ranging from 0.01 m/s to 30 m/s. Lognormal distribution requires three parameters and the parameter values are listed accordingly. Figure 1 shows the CDF of raw data and that of Lognormal distribution. Since these 20 distributions have statistic values lower than 0.1, they are considered a relatively good fit for practical purposes. The graphs show similar results and the similarity is proven by their statistic values.

TABLE II. Parameters for Distributions of Speeds
from 0.01 m/s to 30.00 m/s.

0.01 - 30.00			
Rank	**Distribution**	**Statistic**	**Parameter**
1	Lognormal	0.03859	σ = 1.4072, μ = 0.89733, γ = 0
2	Fatigue Life(3P)	0.04272	α = 1.3456, β = 2.8829, γ = -0.18893
3	Lognormal(3P)	0.04303	σ = 1.2761, β = 0.98033, γ = -0.07749
4	Log-Logistic(3P)	0.04461	α = 1.2533, β = 2.5724, γ = 6.7623E-4
5	Phased Bi-Weibull	0.04772	$\alpha1$ = 1.06, $\beta1$ = 3.619, $\gamma1$ = 0, $\alpha2$ = 0.72755, $\beta2$ = 4.2933, $\gamma2$ = 2.49
6	Frechet(3P)	0.05133	α = 1.1741, β = 2.213, γ = -0.59554
7	Dagum	0.05244	κ = 0.69969, α = 1.441, β = 3.8884, γ = 0
8	Johnson SB	0.05525	γ = 1.2294, δ = 0.50855, λ = 28.515, ξ = 0.21335
9	Dagum(4P)	0.05645	κ = 0.63721, α = 1.4529, β = 4.3813, γ = 0.01
10	Log-Pearson 3	0.05663	α = 11.523, β = -0.41454, γ = 5.6741
11	Pearson 6(4P)	0.06449	$\alpha1$ = 0.87855, $\alpha2$ = 3.8949, β = 17.718, γ = 0.01
12	Burr(4P)	0.06806	κ = 6.9702, α = 0.89148, β = 36.702, γ = 0.01
13	Gen. Gamma(4P)	0.06837	κ = 0.75574, α = 1.1357, β = 3.8121, γ = 0.01
14	Gen. Pareto	0.06908	κ = 0.33377, σ = 3.6824, μ = 0.01
15	Pareto 2	0.06970	α = 3.0493, β = 11.325
16	Pearson 6	0.07269	$\alpha1$ = 0.93678, $\alpha2$ = 3.6704, β = 15.511, γ = 0
17	Weibull(3P)	0.07308	α = 0.817585, β = 4.7174, γ = 0.01
18	Burr	0.07552	κ = 13.116, α = 0.86664, β = 87.161, γ = 0
19	Gamma	0.07870	α = 0.71417, β = 7.4253, γ = 0
20	Weibull	0.07929	α = 0.83261, β = 4.7886, γ = 0

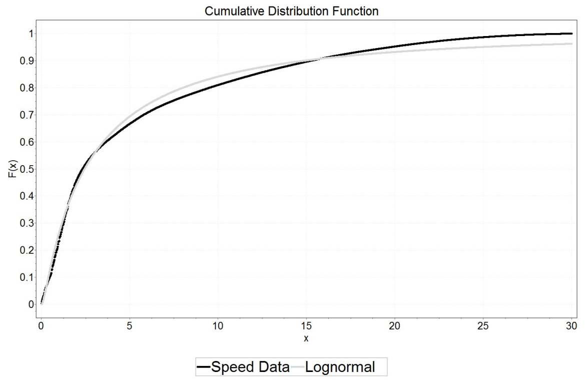

Figure 1. Lognormal Distribution from 0.01 m/s to 30.00 m/s.

TABLE III. Parameters for Distribution of Speeds from 0.01 m/s to 50.00 m/s.

0.01 - 50.00			
Rank	**Distribution**	**Statistic**	**Parameter**
1	Lognormal	0.03669	σ = 1.4235, μ = 0.92117, γ = 0
2	Log-Logistic(3P)	0.04015	α = 1.2355, β = 2.6226, γ = 0.00216
3	Lognormal(3P)	0.04172	σ = 1.3013, μ = 0.99766, γ = -0.07077
4	Fatigue Life(3P)	0.04462	α = 1.3809, β = 2.9546, γ = -0.18301
5	Phased Bi-Weibull	0.04706	$\alpha1$ = 1.0588, $\beta1$ = 3.6587, $\gamma1$ = 0, $\alpha2$ = 0.74503, $\beta2$ = 4.3098, $\gamma2$ = 2.48
6	Pearson 6(4P)	0.04929	$\alpha1$ = 0.93411, $\alpha2$ = 2.7027, β = 10.366, γ = 0.01
7	Frechet(3P)	0.05082	α = 1.1418, β = 2.2015, γ = -0.56667
8	Dagum(4P)	0.05108	κ = 0.62862, α = 1.4854, β = 4.3338, γ = 0.01
9	Dagum	0.05157	κ = 0.72888, α = 1.3983, β = 3.8113, γ = 0
10	Log-Pearson 3	0.05609	α = 13.024, β = -0.39444, γ = 6.0584
11	Burr(4P)	0.05750	κ = 5.6601, α = 0.8748, β = 29.514, γ = 0.01
12	Gen. Pareto	0.06641	κ = 0.37942, σ = 3.6578, μ = 0.01
13	Pareto 2	0.06701	α = 2.6758, β = 9.8706
14	Pearson 6	0.06824	$\alpha1$ = 0.96916, $\alpha2$ = 2.8621, β = 11.213, γ = 0
15	Burr	0.07062	κ = 5.3507, α = 0.89947, β = 27.138, γ = 0
16	Weibull(3P)	0.07084	α = 0.7923, β = 4.8503, γ = 0.01
17	Gamma	0.07284	α = 0.65148, β = 8.5636, γ = 0
18	Weibull	0.07926	α = 0.81451, β = 4.9539, γ = 0
19	Gen. Gamma(4P)	0.08034	κ = 0.89068, α = 0.82445, β = 6.6063, γ = 0.01
20	Gamma(3P)	0.07084	α = 0.70851, β = 7.8447, γ = 0.01

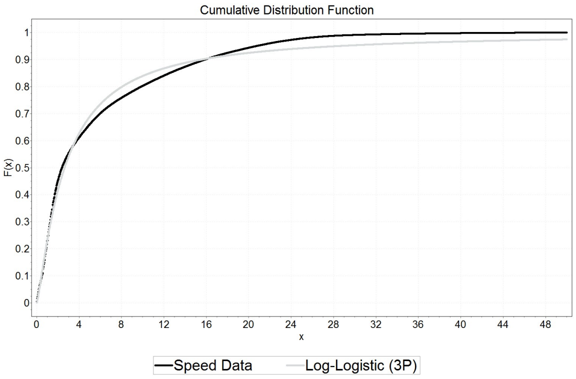

Figure 2. Log-Logistic(3P) Distribution from 0.01 m/s to 50.00 m/s.

Table III summarizes 20 distinguished distributions for the speed range of 0.01 m/s to 50 m/s. Again, lognormal shows the highest rank along with log-logistic(3P) and lognormal (3P) distribution with even smaller statistic values.

It is clear that these distributions must have different parameters than in the case for the speed range of 0.01 m/s to 30 m/s. That is, there are slight shifts in the parameters of the distributions. However we successfully deduced the common probability distribution for human mobility speeds. Figure 2 shows CDF of raw data versus CDF of Log-Logistic (3P) in the speed range of 0.01 m/s - 50.00 m/s. Phased Bi-Weibull distribution is composed of two independent Weibull Distribution over two disjoint domain.

Table IV summarizes 20 distinguished distributions for the speed range of 2.78m/s to 30m/s. Distributions such as Kumaraswamy as shown in Figure 3 is the most distinguished one.

TABLE IV. Parameters for Distributions of Speeds from 2.78 m/s to 30.00 m/s.

2.78 - 30.00			
Rank	**Distribution**	**Statistic**	**Parameter**
1	Kumaraswamy	0.01883	$\alpha1$ = 0.77339, $\alpha2$ = 2.0933, a = 2.78, b = 31.475
2	Gen. Gamma(4P)	0.02436	κ = 2.0551, α = 0.34749, β = 16.827, γ = 2.78
3	Beta	0.02918	$\alpha1$ = 0.67033, $\alpha2$ = 1.8405, a = 2.78, b = 30.349
4	johnson SB	0.03251	γ = 0.87714, δ = 0.67248, λ = 27.818, ξ = 2.3027
5	Gamma(3P)	0.03949	α = 0.93084, β = 8.0734, γ = 2.78
6	Weibull(3P)	0.04194	α = 0.99716, β = 7.4181, γ = 2.78
7	Burr(4P)	0.04217	κ = 9.3570E+5, α = 1.0387, β = 4.1967E+6, γ = 2.7778
8	Exponential(2P)	0.04502	λ = 0.13589, γ = 2.78
9	Erlang(3P)	0.04509	m = 1, β = 7.3576, γ = 2.78
10	Pearson 6(4P)	0.05484	$\alpha1$ = 0.98461, $\alpha2$ = 403.79, β = 2933.8, γ = 2.78
11	Fatigue Life	0.05600	α = 0.66622, β = 8.3005, γ = 0
12	Log-Pearson 3	0.05799	α = 918.16, β = 0.02129, γ = -17.436
13	Lognormal	0.05964	σ = 1.3013, μ = 0.99766, γ = -0.07077
14	Gen. Pareto	0.06220	κ = -0.25534, σ = 9.3381, μ = 2.78
15	Lognormal(3P)	0.06696	σ = 0.95749, μ = 1.6937, γ = 2.1059
16	Pearson 6	0.06816	$\alpha1$ = 43.735, $\alpha2$ = 2.8405, β = 0.45116, γ = 0
17	Pearson 5(3P)	0.06930	α = 2.7254, β = 18.727, γ = -0.05896
18	Pearson 5	0.06977	α = 2.6836, β = 18.223, γ = 0
19	Reciprocal	0.06991	a = 2.78, b = 30.0
20	Frechet(3P)	0.07047	α = 2.2951, β = 8.3746, γ = -2.1033

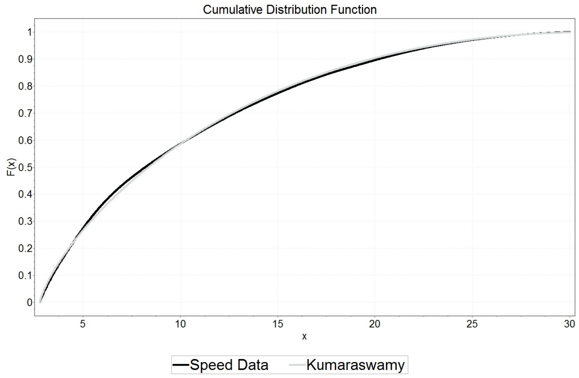

Figure 3. Kumaraswamy Distribution from 2.78 m/s to 30.00 m/s.

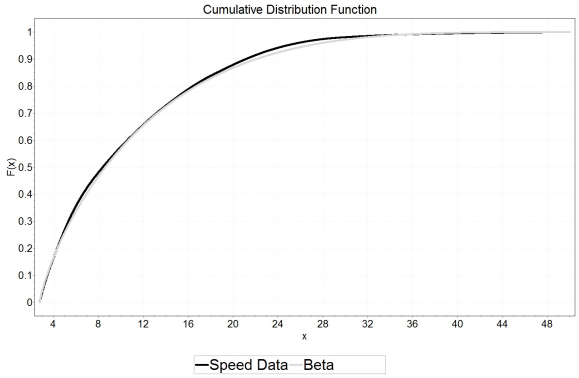

Figure 4. Beta Distribution from 2.78 m/s to 50.00 m/s.

Table V summarizes 20 distinguished distributions for the speed range of 2.78 m/s to 50 m/s. Distributions such as Beta as shown in Figure 4, Weibull(3P) as shown in Figure 5 and

Burr(4P) were the best tree distributions in this speed category.

TABLE V. Parameters for Distributions of Speeds from 2.78 m/s to 50.00 m/s

2.78 - 50.00			
Rank	**Distribution**	**Statistic**	**Parameter**
1	Beta	0.02177	$\alpha1$ = 0.77961, $\alpha2$ = 3.9036, a = 2.78, b = 51.667
2	Weibull(3P)	0.03077	α = 0.96766, β = 7.8915, γ = 2.78
3	Burr(4P)	0.03243	κ = 7.5068E+8, α = 1.0113, β = 4.7291E+9, γ = 2.7796
4	Exponential(2P)	0.03411	λ = 0.12722, γ = 2.78
5	Erlang(3P)	0.03424	m = 1, β = 7.8579, γ = 2.78
6	Gen. Pareto	0.03703	κ = -0.07541, σ = 8.3914, μ = 2.78
7	Gen. Gamma(4P)	0.03925	κ = 0.99856, α = 0.9238, β = 8.4445, γ = 2.78
8	Pearson 6(4P)	0.04931	$\alpha1$ = 0.91756, $\alpha2$ = 20.15, β = 161.89, γ = 2.78
9	Gamma (3P)	0.05284	α = 0.91908, β = 8.1668, γ = 2.78
10	Log-Pearson 3	0.05538	α = 205.5, β = 0.04676, γ = -7.4669
11	Fatigue Life	0.05690	α = 0.69581, β = 8.5747, γ = 0
12	Lognormal	0.05941	σ = 0.67026, μ = 2.1414, γ = 0
13	Lognormal(3P)	0.06369	σ = 1.0099, μ = 1.6995, γ = 2.2037
14	Log-Logistic(3P)	0.06389	α = 1.4469, β = 5.1199, γ = 2.6294
15	Pearson 6	0.06416	$\alpha1$ = 86.972, $\alpha2$ = 2.5986, β = 0.20853, γ = 0
16	Burr	0.06526	κ = 1.2717, α = 2.2942, β = 9.8812, γ = 0
17	Pearson 5	0.06531	α = 2.5322, β = 17.465, γ = 0
18	Fatigue Life(3P)	0.06538	α = 1.0624, β = 5.4173, γ = 2.0879
19	Log-Gamma	0.06617	α = 10.207, β = 0.20979
20	Pearson 5(3P)	0.06718	α = 2.4103, β = 15.982, γ = 0.18981

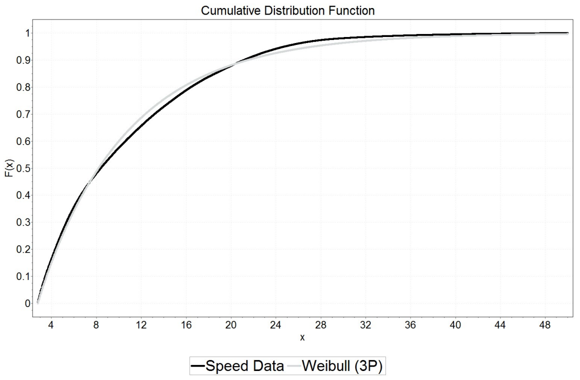

Figure 5. Weibull(3P) Distribution from 2.78 m/s to 50.00 m/s.

IV. CONCLUSIONS AND FUTURE WORK

In this research, we showed several possible probability distributions of speed. Here, speed is used to mean the possible speed found within everyday human life. Using the positioning data sets, speed values were calculated. These positioning data sets were collected by the use of mobile positioning devices such as GPS receivers or smartphones. Volunteers carried such devices in order to collect positioning data.

It is normal that people stay at a certain place for a while; thus, our data included many speed values of zero and there was a zero inflated probability distribution. Using the unit of 0.01 m/s, we divided the range of speeds in four groups. For each category, we executed the Kolmogorov-Smirnov test to find an acceptable approximation of probability distribution. As a result, we provided several well-fit probability distributions for speed. Different from our previous research, we found better fits through using this more precise unit and with more data. One of the notable distributions is exponential distribution, which is ranked 4th in Table V in the speed range of (7.2 Km/h, 180.0 Km/h). Since it has pretty nice statistic, the exponential distribution could be used as an alternative for less strict applications for the devices with lower computational power than other complicated distributions.

We expect that this basic research will help other researchers develop or assess location based services, mobile computing, positioning devices, and others. For example, detection of positioning error is a likely use of our findings. It is well known that positioning data develops errors, which are in fact mostly due to systematic and environmental errors. More precisely, errors in positioning data in the form of (latitude, longitude) follow bivariate normal distribution. Therefore, the speed values derived from positioning data also show propagated errors. It is very hard for mobile devices to detect positioning errors and the derived errors since a user can rarely touch the underlying positioning system or change the operating environment.

One possible scenario can be described: once we have a sequence of speed values of 5.1, 5.2, 5.2, 5.3, 11.8 and 5.3 we can easily identity that 11.8 is the abrupt change of speed value and it may imply an error on the positioning tuple related to the speed value. In such a case, we need to develop sophisticated method to determine the *abruptness* of speed change and we guess that the method may be based on a statistical method. The results shown in this paper may thus be a basis for developing positioning error detection as shown in [16] where it was assumed that human mobile speed is up to normal distribution. However, more precise distribution such as lognormal distribution and beta distribution can be used as basis of erroneous positioning data detection and moreover exponential distribution can be used as a real-time application of this approach since exponential distribution is simple enough to be used on mobile devices.

V. ACKNOWLEDGMENTS

This research was supported by the Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (NRF-2012R1A2A2A03046473).

REFERENCES

[1] “iOS 7: Understanding location services,” http://support.apple.com/kb/HT5594 [retrived: Jul, 2014].

[2] C. Robusto, “The cosine-haversine formula,” American Mathematical Monthly, 1957, pp. 38–40.

[3] M. C. Gonzalez, C. A. Hidalgo, and A-L Barabasi, “Understanding individual human mobility patterns,” Nature, vol. 453, no. 7196, 2008, pp. 779–782.

[4] M. Kim, K. David and S. Kim, “Extracting a mobility model from real user traces.” in INFOCOM, vol. 6, 2006, pp. 1–13.

[5] L. Xiao, Z. Xudong, L. Weifeng, Z. Tongyu, and X. Ke, “The scaling of human mobility by taxis is exponential,” Physica A: Statistical Mechanics and its Applications, vol. 391, no. 5, 2012, pp. 2135–2144.

[6] K.W. On, and H.Y. Song, “On the probability distribution of speeds derived from positioning data obtained by portable positioning devices,” in Proceedings of International Conference on Advances in Mobile Computing and Multimedia. New York, NY, USA: ACM, 2013, pp. 316–320.

[7] F. J. Massey Jr, “The kolmogorov-smirnov test for goodness of fit,” Journal of the American statistical Association, vol. 46, no. 253, 1951, pp. 68–78.

[8] "iPhone 4," http://support.apple.com/kb/sp587 [retrived: Jul, 2014].

[9] "iPhone 4s," http://www.apple.com/iphone-4s/specs/ [retrived: Jul, 2014].

[10] "Galaxy s3," http://www.samsung.com/global/galaxys3/ [retrived: Jul, 2014].

[11] "Galaxy note2," http://www.samsung.com/global/microsite/galaxynote/note2/index.html?type=find [retrived: Jul, 2014].

[12] "Garmin gpsmap800," https://buy.garmin.com/en-US/US/into-sports/cycling/edge-800/prod69043.html [retrived: Jul, 2014].

[13] "Garmin gpsmap810," https://buy.garmin.com/en-US/US/into-sports/cycling/edge-810/prod112912.html [retrived: Jul, 2014].

[14] "Garmin gpsmap62s," https://buy.garmin.com/shop/shop.do?pID=63801 [retrived: Jul, 2014].

[15] F. Scholz, "Maximum likelihood estimation," Encyclopedia of Statistical Sciences, 1985.

[16] K.W. On, and H.Y. Song, "Detection and estimation of erroneous positioning data," Procedia Computer Science, vol. 18, 2013, pp. 2533–2536.