Travel Spend Error Estimation Method based on Handoff Data

Mengyu Rao[1,2], Gaosheng Lin[1], Weiping Chen[1]*
[1]Fuzhou University, Fuzhou, China
[2]National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland
831903230@fzu.edu.cn, 200327170@fzu.edu.cn, w.p.chen@fzu.edu.cn
*Corresponding author

Abstract—**With the development of cities and the continuous progress of people's lives, more and more cities have become congested. Many methods of estimating vehicle speed have been proposed for travel convenience, but these methods have large errors and low accuracy. Currently, a method for evaluating the error of estimated vehicle speed is highly needed. A method for estimating handoff-based travel spend error is highlighted in this paper. The result shows that the probability distribution of handoff locations errors is closed relative to normal distribution. Given the data, the mean squared error (MSE) of the estimated travel distance between two locations can be quickly calculated. Moreover, the relation between the cumulative distribution function (CDF) and the cumulative distribution function of normal distribution (CDFND) is in a 98% confidence interval.**

Keywords—Handoff data, Telecommunication, Travel Spend Error

I. Introduction

With the development of telecommunication and communication technologies, mobile network data plays a significant role in fetching traffic information, such as travel speed estimation, traffic congestion conditions, and etc [1].

Many previous studies show that the handoff data is a critical element of travel speed and travel time estimation. Gundlegard et al. in [2][3] concluded that the travel time in an urban environment could be estimated by the handoff data in GSM and UMTS. In [4][5], travel time was examined by Janecek et al. using 2G/3G network signaling data.

Recently, some handoff-based travel speed estimation methods have been developed. For instance, Meniem et al. in [6] developed a method for speed estimation and map matching using handover among Base Transceiver Stations (BTS), and Yang et al. in [7] examined a travel time estimation model for urban arterials based on Hellinga et al.'s model [8] by using the handoff data in Chengdu, China.

Handoff-based travel speed estimation methods have been posed from 2G to 4G. However, there are a few methods for analyzing handoff-based travel speed estimation errors at present. Handoff-based travel speed estimation errors present a challenge to these travel speed estimation methods. Large errors will cause the results of these methods to be inaccurate and will have a specific impact. This paper innovatively proposes a method to estimate handoff-based travel spend error.

II. Related work

The paper intersects the methods of travel speed estimation errors field in telecommunication and communication technologies. The following paragraphs cover related papers.

A. Using GSM Data to Estimate Travel Speed Errors

Interest in travel speed estimation errors has rapidly increased in the last two decades. The first study in this direction was published in 1997 [9]. This study aimed to estimate travel speed and time using GSM data. However, estimating with this method has a large error. In 2000, Ygnace et al. [10] put forward a method with low positioning error constant speed by Gaussian distribution. Shortly afterward, according to the SACCH signals from the mobile, Romy [11] successfully estimated the positions using GSM data in 2001. In 2002 [12], a pilot service used GSM network data to monitor travel time developed by the Finnish Road Administration (Finnra) in cooperation with a private sector. They posed a possibility to detect whether there was a traffic accident or not. However, an unfixed position may lead to inaccurate results. Hsiao et al. [13] put forward an estimation method based on segments instead of distance. Under normal conditions, this method can significantly improve the accuracy compared with distance-based methods.

B. Using CDR Data to Estimate Travel Speed Errors

According to [14][15], CDR data can provide more information. One of the studies that opened up this direction was carried out by Wang et al. [16]. In 2016, Kujala et al. [17] collected statistical properties of users' trajectories and estimated travel time city to city between 62 cities of Senegal. However, there was only a slightly linear dependence between them compared with Google's API. Several years later, Hasan et al. [18] post a transient travel extraction method between TBSs. Another method which was put forward by Batran et al. [19] focused on OD pairs' average travel time. In 2019, Mai et al. [20] put a method and estimated travel time in Guangzhou City. Though the method is similar to [18], this method has some innovative features and achieved high accuracy of 67%.

In contrast, a simple method of travel speed estimation errors is proposed in this paper. The evaluation of this method is in Section IV.

III. Method and Analysis

In this section, the idea that the probability distribution of handoff locations errors is a normal distribution is proposed. The method of Handoff-based travel speed estimation errors is shown in Section A, and the analysis of Handoff-based travel speed estimation errors is shown in Section B.

A. Method

This paper proposes a method (shown in Figure 1) that the handoff from Cell $(i-1)$ to Cell (i) is performed by OBU j at location $l_{i,j}$ at time $t_{i,j}$, and another handoff from Cell (i) to Cell $(i+1)$ is performed by OBU j at location $l_{i+1,j}$ at time $t_{i+1,j}$. The time difference between time $t_{i,j}$ and time $t_{i+1,j}$ is assumed as $T_{i,j}$. Therefore, the practical travel distance could be measured as $\left(l_{i+1,j}-l_{i,j}\right)$, and the practical travel speed $U_{i,j}$ could be measured by Equation .

$$U_{i,j}=\frac{l_{i+1,j}-l_{i,j}}{t_{i+1,j}-t_{i,j}}=\frac{l_{i+1,j}-l_{i,j}}{T_{i,j}} \quad (1)$$

The mean location of handoff in the historical dataset from Cell$(i-1)$ to Cell (i) is l_i, and the mean location of handoff in the historical data set from Cell (i) to Cell $(i+1)$ is l_{i+1}. Therefore, the estimated travel distance could be assumed as $(l_{i+1}-l_i)$, and the estimated travel speed of OBU j, $u_{i,j}$ could be calculated by Equation (2).

$$u_{i,j}=\frac{l_{i+1}-l_i}{t_{i+1,j}-t_{i,j}}=\frac{l_{i+1}-l_i}{T_{i,j}} \quad (2)$$

The estimated travel distance from location l_i to location l_{i+1}could be assumed as d_i.The estimated handoff location from Cell $(i-1)$ to Cell (i) is $\epsilon_{i,j}$ for OBU j, and the estimated handoff location from Cell (i) to Cell $(i+1)$ is $\epsilon_{i+1,j}$ for OBU j. Therefore, the relationship between the practical travel distance $D_{i,j}$ of OBU j and the estimated travel distance d_i is presented as Equation (3).

$$D_{i,j}=l_{i+1,j}-l_{i,j}=d_i+\epsilon_{i,j}+\epsilon_{i+1,j} \quad (3)$$

Therefore, the mean squared error (MSE) of the estimated travel distance between location l_i to location l_{i+1} could be calculated by Equation (4).

$$\sum_{\epsilon_{i,j}=-\infty}^{n}(D_{i,j}-d_i)^2=\sum_{j=1}^{n}(\epsilon_{i+1,j}-\epsilon_{i,j})^2 \quad (4)$$

The probability density function (PDF) of location error $\epsilon_{i,j}$ is assumed as $P_i(\epsilon_{i,j})$ and the PDF of location error $\epsilon_{i+1,j}$ is assumed as $P_{i+1}(\epsilon_{i+1,j})$. Therefore, the MSE of the estimated travel distance between location l_i to location l_{i+1} could be calculated by Equation (5).

$$\int_{\epsilon_{i,j}=-\infty}^{\infty}\int_{\epsilon_{i+1,j}}^{\infty}(\epsilon_{i+1,j}-\epsilon_{i,j})^2P_i(\epsilon_{i,j})P_{i+1}(\epsilon_{i+1,j})d_{\epsilon_{i,j}}d_{\epsilon_{i+1,j}} \quad (5)$$

B. Analysis

This paper proposes that the probability distribution of handoff location errors is a normal distribution. The PDF of handoff location errors is assumed as the PDF of normal distribution. The PDF of location error $\epsilon_{i,j}$ is assumed as $P_N\left(\epsilon_{i,j},\mu_i,\sigma_i\right)$ with a mean value μ_i and a standard deviation σ_i, and the PDF of location error $\epsilon_{i+1,j}$ is assumed as $P_N\left(\epsilon_{i+1,j},\mu_{i+1},\sigma_{i+1}\right)$ with a mean value μ_{i+1} and a standard deviation σ_{i+1}. Therefore, the MSE of the estimated travel distance between location l_i to location l_{i+1} (shown in Equation (7)) could be calculated by Equation (5) and Equation (6).

$$P_N(\epsilon,\mu,\sigma)=\frac{1}{\sigma\sqrt{(2\pi)}}e^{\frac{1}{2}(\frac{\varepsilon-\mu}{\sigma})^2} \quad (6)$$

$$\int_{\epsilon_{i,j}=-\infty}^{\infty}\int_{\epsilon_{i+1,j}}^{\infty}(\epsilon_{i+1,j}-\epsilon_{i,j})^2P_i(\epsilon_{i,j})P_{i+1}(\epsilon_{i+1,j})d_{\epsilon_{i,j}}d_{\epsilon_{i+1,j}} \quad (7)$$
$$=(\mu_1-\mu_2)^2+{\sigma_1}^2+{\sigma_2}^2$$

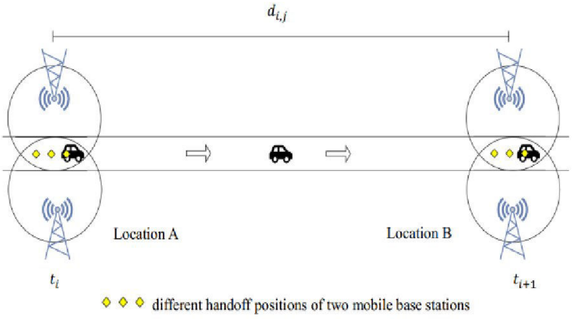

Fig. 1 The scenario of Handoff-based travel form location A to location B.

IV. Practical Experimental Results and Discussion

This section discusses the practical experimental results. The practical experimental environment is shown in Section A, and results are shown in Section B.

A. The Practical Experimental Environment

In the practical experimental environment (shown in Figure 2), the data was collected by a piece of mobile equipment (Huawei P30 running Harmony OS 2.0 with Huawei Kirin 980 processor and the GPS module). An Android application was implemented and used to collect the timestamps, the MNC (Mobile Network Code), the LAC (location area code), the CellID (ID of the coverage area of the mobile phone signal) from mobile base stations, and the coordinates of the GPS module every second (shown in Table I). The handoff points were visualized by Baidu Open Map API. There were 100 records in this experiment, and the number of handoffs was 16515 times. Due to the GPS module error and geographical situation, the identical handoffs happening in every record were not 100%. Therefore, two of the most frequently happening handoffs were selected and grouped. This study selected 32 records, including the GPS module coordinates, the CellID from mobile base stations, and the timestamps. The mean distance between two handoff locations groups is 626.96426m.

TABLE I. Partial Raw Handoff Data

ID	MNC	LAC	CellID	Latitude(°)	Longitude(°)
1	0	24793	107845674	26.0597	119.1864
2	0	24793	107845674	25.0598	119.1881
3	0	24793	107845674	26.0551	119.1882

B. Practical Results and Analysis

Due to the handoff points in every group were not at the exact location, this experiment assumed that these points were in the same line. The line could be calculated by the central positioning coordinates of two handoff groups (shown in Table II and Figure 3).

Fig. 2 The practical experimental environment.

TABLE II. THE AVERAGE LONGITUDE AND LATITUDE OF TWO GROUPS

Handoff Group	Group 1	Group 2
Longitude(°)	119.1882	119.1879
Latitude(°)	26.0591	26.0535

Fig. 3 The central positioning coordinates of two handoff groups.

The corresponding points (α', β') on the line of the actual handoff points could be calculated by the Least Squares. The distance between the central handoff points of the two groups and the corresponding points on the line of the handoff locations points could be calculated by the Spherical Distance Formula.

TABLE III. THE MEAN AND THE STANDARD DEVIATION OF TWO HANDOFF GROUPS

Handoff Group	Group 1	Group 2
Mean	-0.30527705	-0.30636040
Standard deviation	5.15156223	5.88760628

For the data analysis, the mean and the standard deviation of the two handoff groups were calculated (shown in Table III). Therefore, the general normal distribution could be transformed into a standard normal distribution. The actual normal distribution curves of handoff groups (shown in Figure 4 and Figure 5) were similar to the theoretical standard normal distribution curve. The Chi-Square Test was used to test the degree of deviation between the actual normal distribution curve and the theoretical standard normal distribution curve of the two groups. The result showed that the relation between the cumulative distribution function (CDF) and the cumulative distribution function of normal distribution (CDFND) is in a 98% confidence interval (shown in Figure 6 and Figure 7).

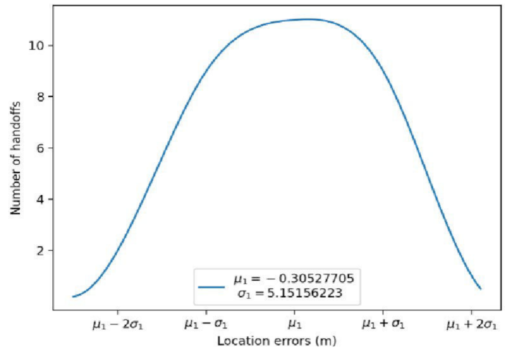

Fig. 4 The normal distribution curve of handoff group 1.

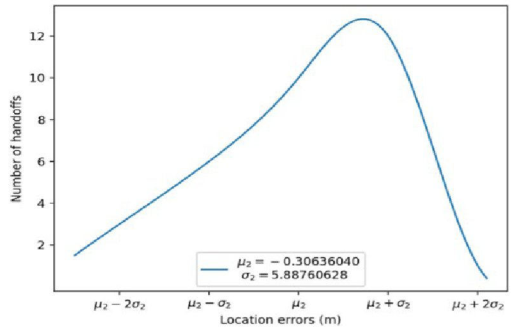

Fig. 5 The normal distribution curve of handoff group 2.

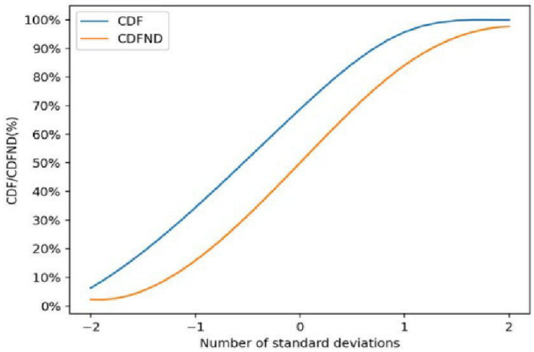

Fig. 6 The cumulative distribution function(CDF) of group 1 and the cumulative distribution function of normal distribution (CDFND) of group 1.

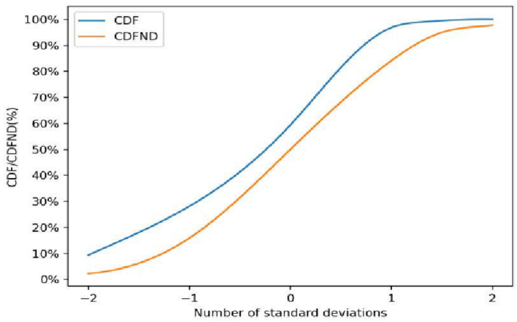

Fig. 7 The cumulative distribution function(CDF) of group 2 and the cumulative distribution function of normal distribution (CDFND) of group 2.

V. CONCLUSIONS

In this paper, the idea that the probability distribution of handoff locations errors is a normal distribution is proposed. The performance of the idea was only verified in Fuzhou University Student Living Area due to the *Covid-2019*. The result showed that the relation between the cumulative distribution function (CDF) and the cumulative distribution function of normal distribution (CDFND) is in a 98% confidence interval.

Although the recent data can confirm the normal distribution of the handoff groups, some challenging problems maybe exist. For instance, less data leads to poor fitting of the normal distribution curve due to the influence of realistic conditions. This can be improved by investigating more data and more mobile equipment.

Furthermore, the proposed idea can be used to analyze and detect road congestion in the future. By collecting the CellID and the handoff data from the routers, the authorities can easily estimate the travel spend between two or more handoff locations. The speed can be used to analyze the road conditions and infer the congestion degree of the road section, providing new solutions for promoting the construction of the Smart City.

ACKNOWLEDGMENT

The authors wish to acknowledge Dr. Chi-Hua Chen, Professor of Fuzhou University, for his help in interpreting the significance of the results of this study. This paper is supported by the 2020 Educational Science Planning Project of Fujian Province (No: FJJKCG20-431), the Undergraduate Education and Teaching Reform Research Project of Fuzhou University (No: School Education [2021] No. 36) and the Innovation and Entrepreneurship Training Program for College Students in Fujian Province (No: S202110386099).

REFERENCES

[1] A. Hadachi and M. Pourmoradnasseri, "A survey on the advancement of travel time estimation using mobile phone network data," IEEE Transactions on Intelligent Transportation Systems, 2021.

[2] D. Gundlegard and J. M. Karlsson, "Handover location accuracy for travel time estimation in gsm and umts," IET Intelligent transport systems, vol. 3, no. 1, pp. 87--94, 2009.

[3] D. Gundlegard and J.M. Karlsson, "Road traffic estimation using cellular network signaling in intelligent transportation systems," 2009.

[4] A. Janecek, K. A. Hummel, D.Valerio, F. Ricciato, and H. Hlavacs, "Cellular data meet vehicular traffic theory: location area updates and cell transitions for travel time estimation," in Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 2012, pp. 361--370.

[5] A. Janecek, D.Valerio, K. A. Hummel, F. Ricciato, and H. Hlavacs, "Thecellular network as a sensor: From mobile phone data to real-time road traffic monitoring," IEEE transactions on intelligent transportation systems, vol. 16, no. 5, pp. 2551--2572, 2015.

[6] M. H. A. Meniem, A. M. Hamad, and E. Shaaban, "Speed estimation and travel direction-detection using handover in gsm: Stability analysis and experimental results in egypt roads," in 2012 8th International Conference on Informatics and Systems (INFOS). IEEE, 2012, pp. CC--7.

[7] F. Yang, Z. Yao, P. J. Jin, and Y. Xiong, "Arterial link travel time estimation considering traffic signal delays using cellular handoff data," IET Intelligent Transport Systems, vol. 13, no. 3, pp. 461--468, 2019.

[8] B. Hellinga, P. Izadpanah, H. Takada, and L. Fu, "Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments," Transportation Research Part C: Emerging Technologies, vol. 16, no. 6, pp. 768--782, 2008.

[9] B. A. N. Mobile et al., "Final evaluation report for the capital-its operational test and demonstration program," 1997.

[10] J.-L. Ygnace, C. Drane, Y. Yim, and R. De Lacvivier, "Travel time estimation on the san francisco bay area network using cellular phones as probes," 2020.

[11] J.-G. Remy, "Computing travel time estimates from gsm signalling messages: the strip project," in ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No. 01TH8585). IEEE, 2001, pp. 6--9.

[12] J. Virtanen, "Mobile phones as probes in travel time monitoring," Finnish Road Administration," 2002.

[13] W. M. Hsiao, S. J. Chang et al., "Segment based traffic information estimation method using cellular network data," in Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, 2005.

[14] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," nature, vol. 453, no. 7196, pp. 779--782, 2008.

[15] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," Nature, vol. 439, no. 7075, pp. 462--465, 2006.

[16] P. Wang, T. Hunter, A. M. Bayen, K. Schechtner, and M. C. Gonzalez, "Understanding road usage patterns in urban areas," Scientific reports, vol. 2, no. 1, pp. 1--6, 2012.

[17] R. Kujala, T. Aledavood, and J. Saramaki, "Estimation and monitoring of city-to-city travel times using call detail records," EPJ Data Science, vol. 5, pp. 1--16, 2016.

[18] M. M. Hasan and M. E. Ali, "Estimating travel time of dhaka city from mobile phone call detail records," in Proceedings of the Ninth International Conference on Information and Communication Technologies and Development, 2017, pp. 1--11.

[19] M. Batran, A. Arai, H. Kanasugi, S. Cumbane, C. Grachane, Y. Sekimoto, and R. Shibasaki, "Urban travel time estimation in greater maputo using mobile phone big data," in 2018 IEEE 20th Conference on Business Informatics(CBI), vol. 2. IEEE, 2018, pp. 122--127.

[20] W. Mai, S. Xie, and X. Chen, "Travel time estimation and urban key routes analysis based on call detail records data: A case study of guangzhou city," in International Conference on Machine Learning and Intelligent Communications. Springer, 2019, pp. 659--676.