International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in; editorijoer@gmail.com

Vol.7., Issue.2, 2019
Mar-Apr

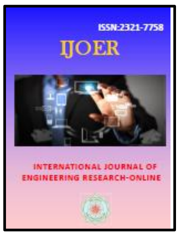
RESEARCH ARTICLE

ISSN: 2321-7758

TRUCK STOP LOCATION ANALYSIS AND ACTIVITY TYPE PREDICTION MODEL BASED ON GPS DATA

R. C. XU*, R.X.ZHONG

Guangdong Key Laboratory of Intelligent Transportation Systems,
School ofIntelligent Systems Engineering, Sun Yat-Sen University, Guangzhou, China.
*Email:xurch3@mail2.sysu.edu.cn
https://doi.org/10.33329/ijoer.72.14



ABSTRACT

Although logistics plays an increasingly important role in transportation planning and decision-making, related studies yet were not sufficient due to the difficulties in height vehicle data availability. With a GPS (Global Positioning System) data set collected from road trucks in Thailand, this paper aimed to obtain the truck stop location by processing and analyzing the raw truck GPS data with a new data processing algorithm. In this study, a robust maximum likelihood estimation method to predict types of activities at truck stops by using passive GPS data was developed. Bayesian-based statistical method updated with known location types was used to enhance the prediction results. Finally, high density truck stop location area could be investigated by spatial clustering algorithm. This research found the proposed raw truck GPS data processing algorithm can accurately obtain the stop location, the accuracy of predicting activity types at GPS stops location based on maximum likelihood estimation and Bayesian method can predict the truck activity type.

I. INTRODUCTION

Height transport has received increasing attention in the traffic planning and decision-making of governments in various countries. Like the passenger traffic activities, there are also various traffic problems that need to be solved. For example, due to the closer cooperation in the Greater Bay Area, China and the improved highway connectivity. The volume of large cargo vehicles such as trucks has increased significantly. It brings tremendous pressure on traffic operations management, traffic safety, road infrastructure, port operations management, and the operation of logistics. In order to strategically and effectively design the logistics infrastructure to meet existing and increasing freight requirements, freight transport analysis is important. However, the main challenge of the above process is lack of the corresponding freight data, such as truck GPS data (including the trajectory and the spatial-temporal information of trucks). With the development of big data, various departments have opened relevant data, it provides a new data environment for relative researches. Therefore, this paper aims to use the collected freight vehicle GPS data to estimate the required freight activity analysis and inferring the fleet management strategies of freight carriers.

To estimate freight analytics, this paper will first develop efficient algorithms for identifying activity stops of freight vehicles from GPS data. Then, based on these identified stops and corresponding characteristics, we will establish machine learning algorithms and trained for estimating the types of activity (e.g., loading) performed at each stop and various freight analytics (e.g., commodity carried) of the tracked vehicles. With the freight analytics of each of the tracked freight vehicles, this paper will develop

KY Publications

algorithms/rules to infer and characterize the fleet management strategies – which are combined sets of trip chaining, vehicle routing, scheduling and vehicle utilization strategies – that freight carriers used to dispatch their freight vehicles. Lastly, the performance of the developed models and algorithms are examined with the use of field data – which are from Thailand.

With the use of GPS data from freight vehicles, this paper has the potential to help transport planners to update the characteristics of freight-related traffic in a more frequent and accurate manner. The proposed algorithms for inferring fleet management strategies of freight carriers help to understand their preference without excessively investigation. This will help planners to propose plans that are generally acceptable by the freight industry in the initial stage of study.

This is particularly important for the planning of logistic infrastructures with rapid changing of freight demands and activity patterns.

II. LITERATURE REVIEW

In the context of intelligent transportation systems (ITS), Hu et al[1]extended the problem of trip table estimations by using adaptive Kalman filtering to estimate the dynamic assignment matrices and OD demands. To consider other traffic model estimations, Lao et al[2]developed a Gaussian mixture model to estimate travel speeds and classified vehicle volumes using loop detectors. Also, Yuan et al[3]adopted the traffic flow model to predict the travel speeds using two traffic data sources (loop detector and floating car data) based on Kalman filtering.

Ideally, the data used in the estimation of OD flows should be collected by GPS-based travel surveys (activity-travel data collected by the GPS equipment attached to probe vehicles or carried by travellers). For instance, Wolf et al[4]developed an automated process to predict travellers' destinations and trip purposes from vehicle GPS traces. In addition, Frignani et al[5] also collected high accurate activity-travel data (e.g. chosen activity type, destination and mode) from internet GPS-based interaction travel feedback systems. To identify the travel path using vehicle GPS traces, a topological map-matching method is normally used (Reference [6]). In addtition, Greaves and Figliozzi[7] use the different of time stamps between GPS-to-satellite communications to determine two major statuses (run or stop). According to their investigations, 240 seconds can be as the acceptable threshold to distinguish the stops from vehicle running states. In addition, McCormack et al[8]found the insignificant truck movements at Washington metropolitan area (e.g. short movements inside large plants or industrial zone) which can be spurious trips. To eliminate such trips, a distance between two consecutive GPS points was adopt. For instance, those trips will be removed if distance between two consecutive GPS points are less than 65 feet. Recently, Yang et al[9]adopted a support vector machine method (SVM) to identify freight delivery stops using GPS data in New York city. The stop features (e.g. stop durations and distances to centre of city) are used to classify all stop into delivery stops and non-delivery stops (e.g. stops due to traffics or rests).

The other method, which can possibly identify activity-travel data, is plate scanning (PS). A comparison of the GPS-based data and plate scanning data reveals that the information obtained from plate scanning is similar to the GPS-based data in the context of tracking vehicles. Based on PS data collection, the information of these vehicles is obtained at pre-determined locations on road network. Additionally, by matching their license plate numbers, the process of plate scanning can identify those vehicles when travelling along a series of plate scanning locations. With this method of data collection, plate scanning is considered to be a method to collect vehicle re-identification (VI) data. The data from plate scanning consists of: (i) the vehicle passing time at plate scanning locations, and (ii) sequence of scanned vehicles along a series of plate scanning locations on road network. The accuracy of collecting the above data from the plate scanning method is determined from detection and identification rates. The detection rate is the proportion of the number of vehicles that pass sensor locations and can be detected (i.e. vehicle is known to have passed a sensor, but the sensor may not be able to identify the vehicle's license plate number). However, a high installation costs of plate scanning sensors may make this approachless applicable than development of GPS-basedsystems for predicting of origin-destination of trucks in the large scale (e.g. whole country of Thailand).

III. OBJECTIVE

In this study, a robust maximum likelihood estimation method to predict types of activities at truck stops by using passive GPS data is developed. The GPS data were obtained from140,000trucks installed with GPS in Thailand. Since each activity types have individual distributions of stop feature, mean and variance of stop durations and the distances of truck stops away from road were used to define activity types at each stop.To enhance the prediction results, Bayesian method updated with known location types was applied.

IV. DATA PROCESS ALGORITHM

A. Truck GPS data set: In this study, the GPS data set collected from 140,000 trucks in Thailand in November 2017. As shown in Fig. 1, the data set includes the identity information (unit id) of the truck, the temporal information (time stamp) of the data, and the spatial information (lat, lon) ofthe trucks in Thailand. The speed information and the vehicle types information of the truck (unit type), wherein the type of the truck includes a private truck, a non-fixed line truck, a fixed line truck, and the like.

	time_stamp	unit_id	lat	lon	speed	unit_type
0	2017-11-03 21:00:01	00700030000000TN6001252589M	13.657573	100.679747	0	1.0
1	2017-11-03 21:00:17	039000100000000000000EE018860	13.882568	100.584500	34	3.0
2	2017-11-03 21:00:32	0290002-201703-00363-000044	14.533698	101.001070	0	6.0
3	2017-11-03 21:00:32	039000100000000000000EE000471	13.758203	100.522844	46	3.0
4	2017-11-03 21:00:16	0290003-201708-00053-000041	14.622970	103.903360	83	1.0

Fig.1 Raw truck GPS data of Trucks

B. ***Clearing and sorting raw GPS data**:* To have effective analysis from raw GPS data, the trucks with incomplete GPS data are removed (e.g. missing GPS time logs or large time gap between GPS records). In addition, raw GPS data of each trucks needs to be sorted into time sequent for determining the distance and difference of time stamps between any two consecutive GPS data points.

C. Identifying potential truck stops: The potential truck stops from raw GPS data are defined by the locations where the trucks stop and do some activities excluding the delay from traffics. For instance, the drivers may be stopping their vehicles for rest at truck parks, rest at way side stops or loading-unloading of the goods.

Generally, trucks tops may be determined from stationary status of two GPS consecutive time stamps. However, some identified stops may be caused by traffic delays. To eliminate these unwanted truck stops (e.g. traffic stops), the minimum stop durations were set. Form previous study, 30minutes dwell-time buffer was used to determine the potential truck stops. Nevertheless, some wanted activity durations in this study are less than 30 minutes. Also, multi-stops are occurred from insignificant movement in some defined areas (e.g. industrial estate, truck terminal, and container yard).

To identify the effective truck stops, it appears that:

1) Stop duration-sd: 20 minutes dwell-time buffer was set to make good balance between elimination of unwanted activities and availability of wanted activities.
2) Travel speed-ts: Average speed between two consecutive GPS data for the same vehicle ID (8km/h).
3) Travel distance-td: Travel distance between two consecutive GPS data for the same vehicle ID (2km).

In this study, we set the three thresholds for the above three judgment indicators. For the stop duration, considering that the truck may be in the service area to rest or queue up for loading, we set a 20-minute threshold to determine whether the stop is a stoppage caused by traffic delays. When the stop duration is less than 20 minutes, we will remove it. When the truck loads or unloads cargo in the parking area or enters the rest area, the speed is small, and there is a certain travel distance. Therefore, we set the average speed and travel distance thresholds is 8 km/h and 2km, respectively. When the data meets the stop duration indicator, it is continued to determine whether the speed between two consecutive GPS data is less than 8 km/h and whether the spatial distance is less than2km. When the three judgment indicators are satisfied, we retain the data to be the stop point data. Therefore, we summarize the above steps to determine the potential stops of trucks in the following data processing algorithm flow chart (Fig. 2).

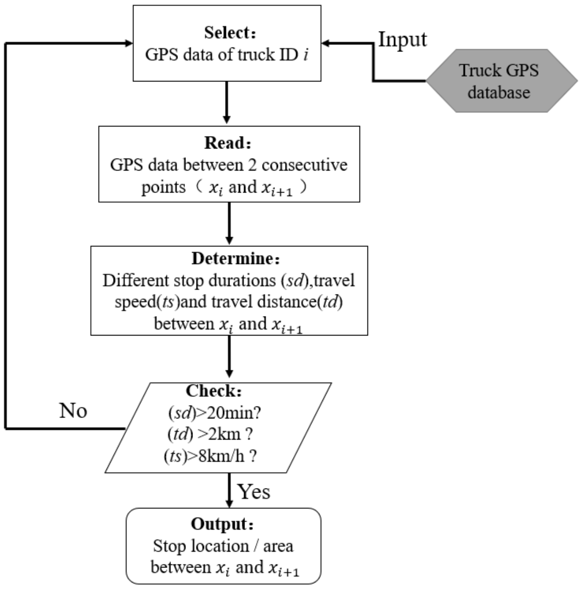


Fig.2 Algorithm for identifying potential truck stops

V. PREDICTING ACTIVITY TYPE

The choice sets of activity type f of travellers are assumed to be given (i.e. $f\in\{1,\ldots,F\}$),in this study, where F is the number of feasible activity types derived from stopped point s. Consider user (truck)i stopping on point s and performing activity type f, the indicator variables can then be observed as follows:

$$z^f_{i,s}=\begin{cases}1, & \text{if user } i \text{, is sequentially stopped} \\ & \text{by point } 1,\ldots,s,\ldots S \\ 0, & \text{otherwise}\end{cases} \tag{1}$$

According to practical distributions of stop features, both vehicle dwell time (y) and the distance (x) from nearest road to stop points are log-normally and independently distributed as follows:

$$\log y_f \sim N\left(E\left(y_f\right),\sigma^2_{y,f}\right) \tag{2}$$

$$\log x_f \sim N\left(E\left(x_f\right),\sigma^2_{x,f}\right) \tag{3}$$

where,

y_f= random dwell times of users doing activity type f.

x_f= random shortest distance from stop points to road of users doing activity type f .

E(y_f), $\delta_{y,f}$= mean and standard deviation of $y_f, y_f\in\mathbf{y}$ and$\delta_{y,f}\in\boldsymbol{\delta}$

E(x_f), $\delta_{x,f}$ = mean and standard deviation of $x_f, x_f\in\mathbf{x}$ and $\delta_{x,f}\in\boldsymbol{\delta}$

To update the model parameter, rewritten now in a vector form **C**={**μ**,**δ**}, based on maximum- likelihood estimation approach, a joint probability density function (log normal distribution) for the random variables(x, y) in (2) and (3) reproducing the observations (**x**,**y**,**z**) is able to be written down as follows:

$$q(\Lambda\,|\,x,y,z)=\prod_{i=1}^{N}\prod_{s=1}^{S}\prod_{f=1}^{F}\Big(f\big(E(y),\sigma_y\,|\,y\big)\cdot g\big(E(x),\sigma_x\,|\,x\big)\Big)^{z^f_{i,s}} \tag{4}$$

where

$$f\big(E(y),\sigma_y\,|\,y\big)=\left(\frac{1}{y_{i,s}\sigma_{y,f}\sqrt{2\pi}}\exp\left(\frac{-\left(\ln y_{i,s}-E(y)\right)^2}{2\sigma^2_{y,f}}\right)\right) \tag{5}$$

$$f\big(E(x),\sigma_x\,|\,x\big)=\left(\frac{1}{x_{i,s}\sigma_{x,f}\sqrt{2\pi}}\exp\left(\frac{-\left(\ln x_{i,s}-E(x)\right)^2}{2\sigma^2_{x,f}}\right)\right) \tag{6}$$

By taking log of function (4), it yields the complete-data log-likelihood of:

$$L(\Lambda\,|\,x,y)=\sum_{i=1}^{N}\sum_{s=1}^{S}\sum_{f=1}^{F}z^f_{i,s}\Bigg[-\left(y_{i,s}\sigma_{y,f}+x_{i,s}\sigma_{x,f}\right)\sqrt{2\pi} -\frac{-\left(\ln y_{i,s}-E(y)\right)^2}{2\sigma^2_{y,f}}-\frac{\left(\ln x_{i,s}-E(x)\right)^2}{2\sigma^2_{x,f}}\Bigg] \tag{7}$$

Given **z** (from the GPS data observations), model's attributes (**μ** and **δ**) are estimated from the likelihood problem (4). Then, by standard laws of conditional probabilities, the probability of truck performing at a sequence of stopped points no s for activity type f is defined as follows:

$$\Pr\left(z^f_{i,s}=1\,|\,x,y,\Lambda\right)=\frac{\Pr\left(z^f_{i,s}=1,x,y\,|\,\Lambda\right)}{\Pr\left(x,y\,|\,\Lambda\right)} \tag{8}$$

By the Bayes' rule (and the fact that $z_{fi,s}$ can only be 0 or 1):

$$\Pr\left(x,y\,|\,\Lambda\right)=\sum_{j=0}^{1}\Pr\left(z^f_{i,s}=j,x,y\,|\,\Lambda\right) \tag{9}$$

Putting (8) into (9), it gives:

$$\Pr\left(z^f_{i,s}=1\,|\,x,y,\Lambda\right)=\frac{\Pr\left(z^f_{i,s}=1,x,y\,|\,\Lambda\right)}{\sum_{j=0}^{1}\Pr\left(z^f_{i,s}=j,x,y\,|\,\Lambda\right)} \tag{10}$$

The probability distributions of selecting activity type f associated with observed stop durations (y) and observed distance from stopped points to nearest road (x) required for the numerator and denominator of (10) can be defined by (4) and another-application of Bayes' rule, yielding:

$$\Pr\left(z_{i,s}^{f}=1 \mid x,y,\Lambda\right)=\sum_{\substack{\text{all comninations } z \\ \text{with } z_{i,s}^{f}=j}} q\left(\mathrm{x},\mathrm{y},\mathrm{z} \mid \Lambda\right) \tag{11}$$

Since the decision on activity type f of user i is independent from other users, the combinations **z** in (11) consist of $z_{fi,s}$=1, when user i selects activity type f and $z_{fi,s}$=0 for other cases that activity type f is not selected from user i. Thus, the probability for user i selecting activity type f can be estimated from the following equation:

$$\Pr\left(f\right)=\Pr\left(z_{i,s}^{f}=1 \mid x,y,\Lambda\right)=\frac{w_{i,s}^{f}}{\sum_{f'=1}^{F} w_{i,s}^{f'}} \tag{12}$$

or a simple form of probability of truck doing activity type (f) can be presented. By applying the Bayesian rule, the posterior probability of the user i stopping at point s (given user in area a) for activity type f is shown as follows:

$$\Pr\left(f \mid a\right)=\frac{\Pr\left(a \mid f\right)\cdot\Pr\left(f\right)}{\Pr\left(a\right)} \tag{13}$$

where,

a= defined areas, which have high density of trucks.

Pr (a)= probability of trucks stops and make some activities in area a.

Pr (f)= (prior) probability of trucks stops and make activity type f shown in (12).

Pr (a / f)= probability of trucks stops in area a, given that trucks make activity type f.

In this study, the choice set of activity types in defined area is known. Also, only two major activity types (type 1: rest/wait and type 2: load-unload) is defined in that area. Given specific activity type which user stops, the probability of trucks stops in defined area is presented by:

$$\Pr\left(a \mid f\right)=\frac{N_{a}^{f}}{\sum_{a'=1}^{A} N_{a'}^{f}} \tag{14}$$

where,

N_a^f= number of trucks stopping in area a, given that trucks make activity type f.

By the law of total probability, probability of trucks stops and make some activities in area a, Pr (a) is defined by:

$$\Pr\left(a\right)=\sum_{f'=1}^{F}\Pr\left(a \mid f'\right)\cdot\Pr\left(f'\right) \tag{15}$$

After inputting (15) into (13), it yields:

$$\Pr\left(f \mid a\right)=\frac{\Pr\left(a \mid f\right)\cdot\Pr\left(f\right)}{\sum_{f'=1}^{F}\Pr\left(a \mid f'\right)\cdot\Pr\left(f'\right)} \tag{16}$$

VI. DBSCAN CLUSTERING ALGORITHM

A. *Introduction of algorithm*

In this study, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm is used to cluster the user (truck) i stopping areas. The DBSCAN algorithm is a density-based spatial clustering algorithm that divides regions of sufficient density into clusters and finds clusters of arbitrary shape in a spatial database with noise, which defines the cluster as the largest point of density set.

The DBSCAN algorithm utilizes the concept of density-based clustering, which requires that the number of intrinsic points (points or other spatial objects) contained in a certain region of the cluster space is not less than a given threshold. The significant advantages of the DBSCAN algorithm are fast clustering and efficient processing of noise points and the discovery of spatial clustering of arbitrary shapes. Therefore, for thetrucks GPS datasetin this study, the advantages of this algorithm are applicable to cluster analysis of truck stopping areas.

B. *Comparison with K-means algorithm*

The purpose of the DBSCAN algorithm is to filter low-density areas and find dense sample points. The principle of the algorithm is consistent with the purpose of this study. Based on the trucks GPS data , this study finds and analyses the high-density stops or areas of trucks by using the appropriate clustering algorithm for the trucks GPS data sets processed by the raw data. Different from the traditional convex clustering based on hierarchical clustering and partitioning clustering, the DBSCAN algorithm can find clusters of arbitrary shapes, and has the following advantages compared with the traditional clustering algorithm:

1. Compared with the K-MEANS clustering algorithm, there is no need to input the number of clusters to be divided (this is consistent with the prior knowledge in this

International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in; editorijoer@gmail.com
Vol.7., Issue.2, 2019
Mar-Apr

study, we do not know how many truckstopping areas for the Thailand;

2. Compared with the hierarchical clustering algorithm and the partitioning clustering algorithm, the shape of the cluster in the DBSCAN algorithm is not biased;
3. We can input the parameters of filtering noise in the clustering process with DBSCAN algorithm;

VII. EXPERIMENT AND RESULTS ANALYSIS

A. ***Stop activity prediction:*** Given two choices of activities possibly made at stop point A from trucks no. 1, (f = 1 for rest and f=2 for pick up-delivery purpose), we assumed that:

i. probability of truck doing activity (f =1) and activity (f =2) estimated from observedvariables (*x*, *y*) are equal to 0.5 and 0.5, respectively.
ii. To update activity type predictions from land use data (see Table 1), stop point A is located on the known area type (say, industrial estate area).
iii. Truck no. 1 make a stop (point A) in industrial estate (*a*=1).

TABLE 1: FREQUENCY OF ACTIVITY TYPES CLASSIFIED BY AREA TYPE.

Activity type	Area type (*a*)		Sum
	Industrial estate (*a*=1)	Truck park (*a*=2)	
Rest (*f*=1)	1,000	5,400	6,400
Pick-up/Delivery (*f*=2)	8,000	600	8,600
Sum	9,000	6,000	15,000

By applying the Bayesian rule, the posterior probability of truck no. 1 made activity type 2 (pick up-delivery purpose) at stop point A (*a*=1) are shown as follows:

$$\Pr(f=2\,|\,a=1)=\frac{\Pr(a=1\,|\,f=2)\cdot\Pr(f=2)}{\Pr(a=1\,|\,f'=1)\cdot\Pr(f'=1)+\Pr(a=1\,|\,f'=2)\cdot\Pr(f'=2)}$$

where given prior probabilities Pr (f'=1)=0.5 and Pr (f'=2)=0.5

Form Table 1, given the observed type of activities, the probability of trucks no. 1 stops in defined area (*a*=1) is presented by:

$$\Pr\left(a=1\,|\,f^{1}=1\right)=\frac{1000}{1000+5400}=0.156,$$

and $\Pr\left(a=1\,|\,f^{1}=2\right)=\dfrac{9000}{9000+600}=0.930$

Thus,

$$\Pr(f=2\,|\,a=1)=\frac{0.938\cdot 0.5}{0.156\cdot 0.5+0.938\cdot 0.5}=0.856$$

After activity type prediction model is updated with land use information as described above, the probability of this truck make activity type 2 (pick up-delivery purpose) increases from 0.5 to 0.856. Consequently, this activity prediction model is enhanced, and truck no. 1 would be stopped (at stop A) for making activity type 2. We can organize it into the flowing flow chart.

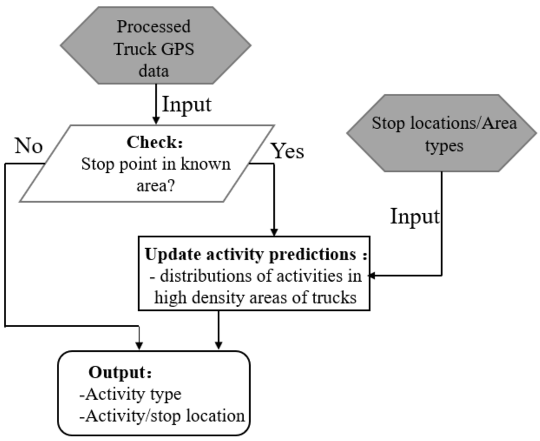


Fig.3Algorithm for identifying truck stops types

B. ***Stop location analysis****:* Before clustering the Thailand truck stops, we visualize all the stop data of the truck after the data processing. As shown in Figure 4, the data of the trucks in the week from November 01, 2017 to November 07, 2017, each of bright yellow dot is the stop for truck. As can be seen from Fig.4, the high-density docking area of the truck is roughly in the capital city of Bangkok and the tourist area of Pattaya.

International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in; editorijoer@gmail.com
Vol.7., Issue.2, 2019
Mar-Apr

Fig.4 Distribution of truck stop data

As shown in Fig.4, we can only initially observe that the high-density truck docking area is Bangkok and Pattaya, and other areas with higher density cannot be directly known from the map. Therefore, we use DBSCAN to perform density clustering on the truck stop data in Fig. 5and find the high-density truck stopping area through density clustering. In this study, the key parameters for DBSCAN are set and selected as follows: the distance between two GPS data points can be calculated according to the latitude and longitude of the earth. Considering that one stopping area will stop multiple trucks, the value of ε is 2 m, Min Pts is 45, respectively, and finally we can obtain the core point of the clustered high-density truck stopping area. According to the density of the stopping area, the stopping areas of different densities are visualized. As shown in Fig. 5, we can observe that the highest density of the stopping area in the data center of the trucks is the capital Bangkok area, and the result can also be observed in the visualization of the dataset that the DBSCAB clustering algorithm used in this study and the selection of algorithm parameters are reasonable.

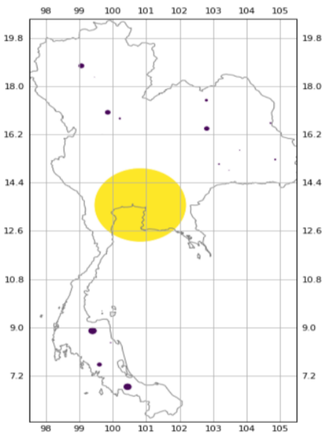


Fig.5 Distribution of truck high density stopping area

C. ***Prediction results analysis:*** We randomly selected approximately 3% of the data set as the test set for this study. As in the previous article, there are still two types of trucks in the test set. Moreover, the parameters of the prediction model are divided into seven types of types of major transport goods in Thailand (i.e., sugar cane, sugar, rice, animal feed, paper, concrete, containers). The mean and standard deviation of the truck stop time and the travel distance in the stopping area in the dataset are obtained by the maximum likelihood estimation algorithm mentioned above. In the experiment, the average stopping time of loading or unloading of trucks in the stopping area is far greater than the rest period, and trucks loaded with paper, containers, and concrete often spend more than 7 hours on loading or unloading in the stopping area. The standard deviation of the waiting time for loading or unloading and rest of trucks in the stopping area is 0.10 and 0.16 respectively. Moreover, for the driving distance of the truck in the parking area, the driving distance of the loading and unloading activities is also larger than the driving distance of the rest. Therefore, this means that the range of loading or unloading areas is large, and trucks often have to be driven directly from the road to a warehouse in the factory for loading or unloading, so the driving distance is large.

VIII. CONCLUSION

This paper has developed a robust maximum likelihood estimation method that use truck GPS data to predict the type of activity of a truck at a stop. The average and variance of the parking duration and the distance travelled by the truck in the docking area are used to calculate or the probability that the vehicle will engage in some type of activity at each stop. In order to evaluate the performance of the proposed method, the estimation results are compared with the conventional methods.

The method proposed in this paper also has a lower error rate in the verification of the test set. Machine learning algorithm (i.e., Support Vector Machine, SVM) is also applicable to the Thai truck GPS dataset used herein. However, more than 40% of the stops and stopping data do not identify the true type of activity of the truck. In order to solve this problem, this paper adopts a Bayesian-based method to update the probability of prediction and improve the prediction accuracy. The use of known land use types in docking areas enhances the accuracy of prediction results.

REFERENCES

[1]. Hu, S.R., Madanat, S., Krogmeier, J., and Peeta, S., 2001. Estimation of dynamic assignment matrices and OD demands using adaptive Kalman filtering .*Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 6(3), 281-300.

[2]. Lao, Y., Zhang, G., Corey, J., and Wang, Y., 2012. Gaussian mixture model-based speed estimation and vehicle classification using single-loop measurements. Journal of Intelligent *Transportation Systems: Technology, Planning, and Operations*, 16(4), 184-196.

[3]. Yuan, Y., Lint, H.V., Wageningen-Kessels, F.V., and Hoogendoorn, S., 2013. Network-widetraffic state estimation using loop detector and floating car data. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*.

[4]. Wolf, J., Schönfelder, S., Samaga, Oliveira, U.M., and Axhausen, K.W., 2004. 80 weeks of GPS-traces: Approaches to enriching the trip information. *Transportation Research Record*,1870, 46-54.

[5]. Frignani, M.Z., Auld, J., Mohammadian, A.K., and Williams, C., 2010. Urban travel route and activity choice survey: Internet-based prompted-recall activity travel survey using global positioning system data. *Transportation Research Record*, 2183, 19-28.

[6]. Felipe, J.; Sergio, M.; Jose, E.N., 2016. Definition of an Enhanced Map-Matching Algorithm for Urban Environments with Poor GNSS Signal Quality. *Sensors*, 2016, 16, 193.

[7]. Greaves, S. P., Figliozzi, M., 2008. Collecting Commercial Vehicle Tour Data with Passive Global Positioning System and Technology. *Transportation Research Record: Journal of the Transportation Research Board*, TRB, National Research Council, Washington, D.C., 158-166.

[8]. McCormack, E. D. et al., 2010. Developing a GPS-Based Truck Freight Performance Measure Platform, *WSDOT research report*, 9-20

[9]. Yang, X., Sun, Xuegang, Z. and Holguín-Veras, J., 2014. Urban Freight Delivery Stop Identification Using GPS Data, *Transport Research Record*, 2411, 55-61.

AUTHOR PHOTOS AND BIOGRAPHY

Ruochen Xu received the B.E. degree in transportation engineering from Hunan City University, China, in 2014. He is currently working toward the M.E. degree in transportation planning and management at Guangdong Key Laboratory of Intelligent Transportation Systems, School of Intelligent Systems Engineering, Sun Yat-Sen University, Guangzhou, China. His research interests include intelligent transportation systems and machine learning and its application.