

Route Choice Behaviour Modeling using IoT Integrated Artificial Intelligence

Joy Iong-Zong Chen,

Department of Electrical Engineering,

Da-Yeh University,

Changhua, Taiwan.

E-mail: jchen@mail.dyu.edu.tw

Jen-Ting Chang

Department of Electrical Engineering,

Da-Yeh University,

Changhua, Taiwan.

Abstract- Automatic Vehicle Identification (AVI) data is used to identify the location of a particular vehicle in and can also be used for route choice behaviour modelling. But the use of AVI doesn't provide accurate information on OD pair and the particular route that is chosen. This problem is addressed in this paper using a semi-supervised learning method which can be used to identify the route on prior training. As the first step, the AVI trace is segregated into observation pairs using the Maximum Likelihood Estimation and then it is further joined with GPS co-ordinates to tackle the sparse issues. As the next step, the heterogeneity and correlation between the various pairs are determined using Mixed Logit model. As the final step, a relationship between the likelihood function and route choice model is established using Maximum to log-likelihood function. Based on the observations, the results are recorded and the proposed work shows significant improvement in the accuracy in route determination. The evaluation scenario shows that the proposed work could be expanded to a larger area. Moreover, the robustness of the system is illustrated using sensitivity analysis. This work uses AVI data with respect to its behaviour in routes through high penetration.

Keywords: Automatic Vehicle Identification, maximum likelihood estimation, GPS Coordination, vehicle location, modelling route choice behaviour

1. Introduction

To identify the route taken by an individual to travel from one place to another, modeling route choice behaviour is essential. This information can be used to understand and analyze the behaviour of the driver, real-time route guidance, traffic control measures and traffic assignment. For applications that require more complex operations, data which is comprehensive and accurate is required. Recent research on route choice behaviour is focused on these aspects [1]. In the beginning, data used for this methodology is collected directly from the traveller in a one-to-one approach. However, as progress in technology grew, this means of gathering information takes place in a number of other manners such as web-based surveys, telephone interviews, questionnaires via mail etc. When data is gathered from the travellers in this manner, it might have many shortcomings like becoming out-of-date quickly, under-reporting, short survey duration, small sampling rate and high collection cost. Advancement in information and communication technologies has led to the possibility of passive data collection from emerging data sources and has proven to be more useful in understanding route choice behaviour of the traveller, enabling better understanding with transportation professionals [2].

In the figure given below, the observations obtained from the GPS gives a reliable and positive estimation of the traveller's choice [3]. The dash lines plotted represent the high-resolution spatiotemporal trajectory. This trajectory will give a clear outline of the trip's whereabouts. Using the observations recorded by the GPS, it is easier to find the right route chosen. There are some limitations that affect the processing and collection of route choice behaviour of the people using GPS data. The major error that is detected often is the position error [4]. Though the GPS coordinates are accurate to a level, there are many instances when their position is thrown off depending on the receiver characteristics and satellite coverage. As indicated in Fig.1, the GPS points tend to go wayward from their actual location resulting in error occurrence. Another hindrance is the size of the sample [5]. A typical set up of the GPS technologies will need power supply and position sensor. The cost of GPS unit will have an impact on analysing samples which are large. The small sample set that is obtained from the GPS will not be sufficient to properly represent the position, especially in the case of buses and taxis [6].

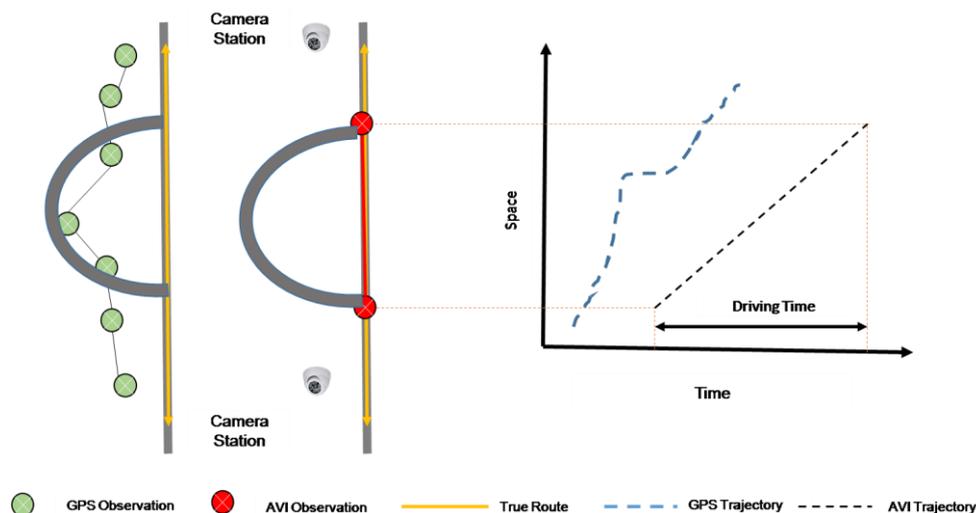


Fig.1. Difference between AVI and GPS data

In [7], the authors examined the shortcomings of the previous approach and proposed the use of detector route choice research progress to gather information on traffic through Automatic Vehicle Identification methodology. It operates such that as the vehicle moves in a certain pattern, it is recorded passively with the help of devices such as camera, radiofrequency identification, transponders, laser and fixed sensor [8]. In comparison to the GPS, the proposed AVI data holds a number of positive attributes, as given below:

1. 24 hours working: One can trace the movement pattern of the vehicles round the clock using AVI whereas a typical GPS will be activated only when the sensor is turned on and will also require more energy supply to work.
2. Observation Accuracy: Using the location information obtained from the sensor, it is possible to determine the accurate coordinates of the vehicle tracked. On the other hand, data provided by the GPS on the vehicle's location will always contain some error. Moreover, it is also possible to calculate the time taken for travelling using the timestamps from nearby AVI observations.
3. Increased Sample Size: The AVI system has the capability of recording the information of all the vehicles that come within a specific radius of the AVI sensor station. Hence, a proper record of all the vehicles that are running on the road will be recorded and their movement pattern is saved in the database.

Using AVI infuses a number of advantages in route choice modelling and also offers feasible solutions to some of the prevalent issues. However, the use of AVI data will not aptly help to review the people's route choice behaviour [9]. From Fig.1, it can be observed that the true choice of route when there are two alternative routes cannot be inferred using AVI observations and identifying the alternative route taken will prove to be tasking. The spatial resolution of the AVI trace also makes it difficult to identify the path taken [10-11]. However, researchers propose the use of Maximum Likelihood Estimation to estimate the required parameters from the route choice model. To implement this, the data on OD pair and specific true chosen route is essential and accordingly, a relationship between the likelihood function and route choice model is established. To overcome these difficulties we propose to use a semi-supervised learning method that can train the route choice model based on the information gathered from AVI [12].

2. Proposed Methodology

2.1 Problem Formulation

A directed graph is used to represent the road network using the equation $D_G = (\{n\}, \{1\})$. Here $\{1\}$ denotes the edges of the road links and $\{n\}$ denotes the nodes representing the terminal points and intersections. For the n^{th} sample trip, can be written as $P(R_{i,n} | \varphi_{c_o \rightarrow c_D}, \alpha)$, representing the probability that a specific route is chosen within the defined based on the difference between the destination (c_D) and origin Traffic Analysis Zone (TAZ). Here α is used to represent the unknown parameter that is to be calculated. The unsupervised learning approach model is also known as the network-free model. IT provides the users with the option to understand and learning

the route choice behaviour of the individuals. The likelihood function of this model, taking into account the data from GPS can be expressed as eq (1).

$$L(\alpha) = \prod_{i=1}^N P(T_n^{GPS}) \quad (1)$$

Where GPS trajectory can be represented as T_n^{GPS} with n sample trips and the probability is expressed as $P(T_n^{GPS})$. There are 3 attributes related with the GPS trajectory namely vehicle ID, timestamp and a coordinate pair.

On decomposing the observed probability with the OD measurement, we can express it using eq (2)

$$P(T_n^{GPS}) = \sum_{c_0, c_D \in \lambda} P(c_0, c_D | \lambda) \sum_{R_{i,n} \rightarrow \varphi_{c_0 \rightarrow c_D}} P(R_{i,n} | \varphi_{c_0 \rightarrow c_D}, \alpha) P(T_n^{GPS} | R_{i,n}) \quad (2)$$

2.2 Semi-Supervised Route Choice

Depending on the AVI trajectory, we can derive the likelihood function through a semi-supervised methodology. The joint probability samples observed together is known as the likelihood function. Thus the trajectory can be expressed based on the chronological order of $\beta_{1,n} \rightarrow, \dots, \beta_{j,n}, \dots, \rightarrow \beta_{j,n}$ where $\beta_{j,n}$ is defined as the j^{th} AVI observation.

$$P(T_n^{AVI}) = \sum_{c_0, c_D \in \lambda} P(C_0 | a_{1,n}) P(C_D | a_{j,n}) P(\beta_{1,n} \rightarrow, \dots, \beta_{j,n}, \dots, \rightarrow \beta_{j,n}) \quad (3)$$

where $P(C_0 | a_{1,n})$ is the actual inference model and λ is the set of candidate OD pairs.

Based on calculations, the likelihood function of the recorded AVI data can be expressed using eq. (4)

$$L(\alpha) = \prod_{i=1}^N P(T_n^{AVI}) \\ = \prod_{i=1}^N (\sum_{c_0, c_D \in \lambda} P(C_0 | a_{1,n}) P(C_D | a_{j,n}) \prod_{i=1}^N P(\beta_{1,n} \rightarrow, \beta_{j,n} | SR_{i,n}) P(SR_{i,n} | \beta_{1,n} \rightarrow, \beta_{2,n}, \dots, \beta_{j,n})) \quad (4)$$

In order to address the sparsity issue, the OD model ($P(C_0 | a_{j,n})$ and $P(C_D | a_{1,n})$) and the measurement equation ($\beta_{1,n} \rightarrow, \beta_{j,n} | SR_{i,n}$) are incorporated using GPS as well as AVI observations.

Using history field-testing GPS data, we can represent the first observation likelihood as shown in eq. (5)

$$P(a_{1,n} | c_0) = \frac{|c_0 \rightarrow a_{1,n}|}{|c_0|} \quad (5)$$

where the values of $|c_0|$ and $|c_0 \rightarrow a_{1,n}|$ can be counted and $|c_0 \rightarrow a_{1,n}|$ represents a total of sample trips. Prior probability of origins can be defined by eq. (6) by using two information sources

$$P(c_0) = P(\rho)P(d_0)$$

where d_0 is the time taken to travel between the observation and candidate origin. c_0 is the candidate origin with POI density ρ .

4. Results and Discussion

Fig.2 shows the scatter plot that is used to determine the POI density based on the relationship between the origin and normalized frequency.

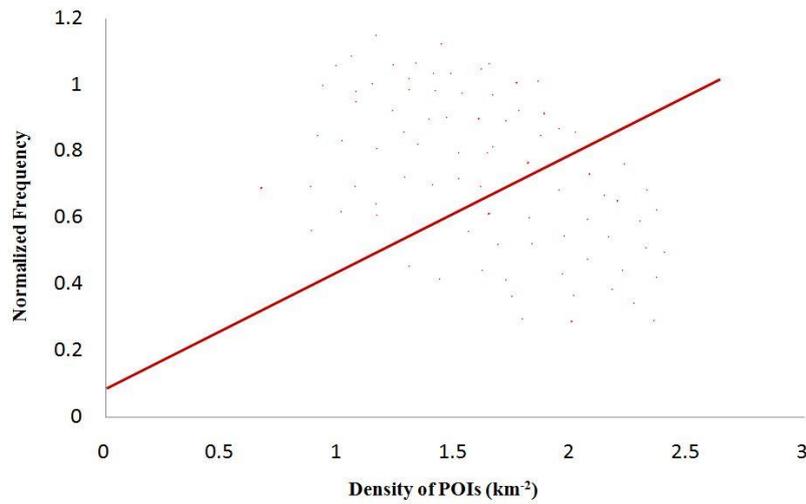


Fig.2. Variation of probabilities with respect to time

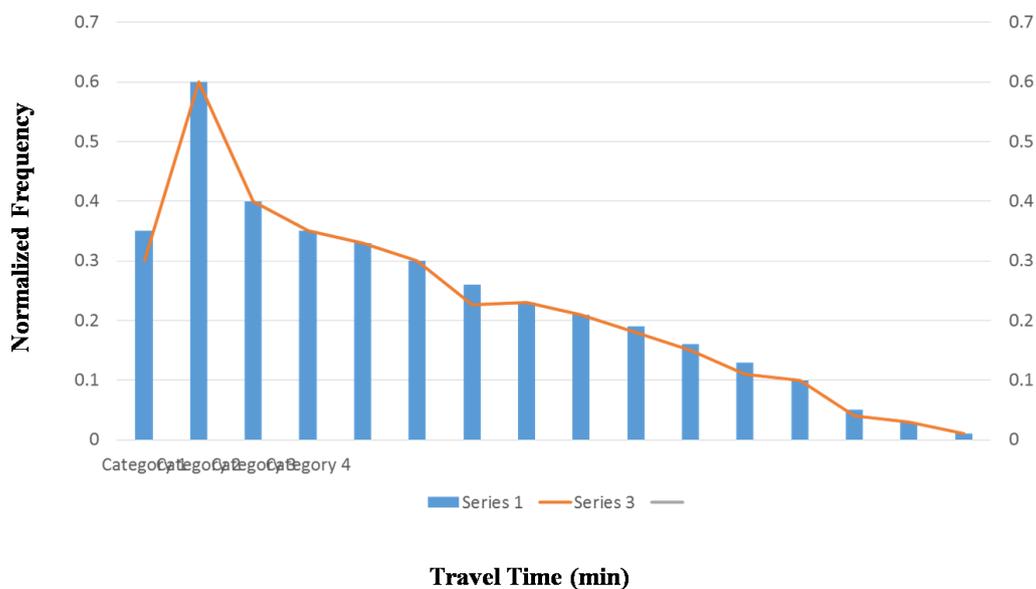


Fig.3. Probability variation with Travel Time

Prior probability that is used to determine the travel time between the first AVI observation and the origin can be represented graphically as shown in Fig.3. There is a high impact of size on the route choice model based on the proposed semi-supervised learning methodology. The average count of alternatives with respect to the alternative generation approach with respect to the varying numbers is represented in bar chart as shown in Fig.4. As the size of the alternative set increases, the accuracy index also increases accordingly. Based on the graph output, it can be identified that it is possible to prevent the identified errors and unnecessary computation burden.

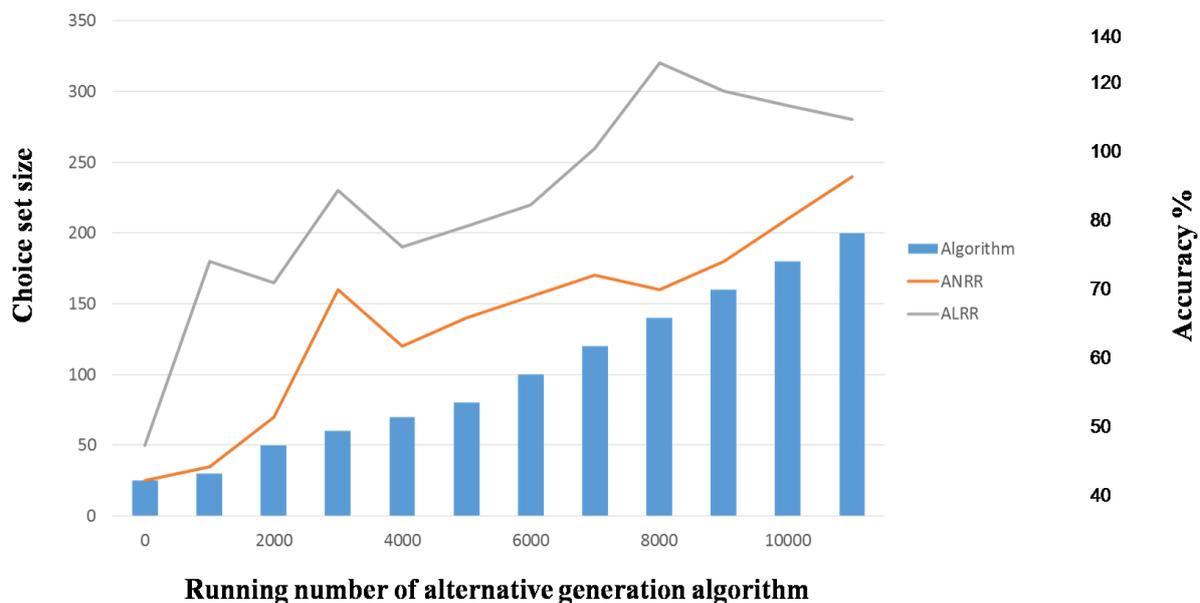


Fig.4. FNC profit

5. Conclusion

The proposed AVI data has a number of advantages such as 24 hours working, accurate observation and large sample size. But the data obtained from AVI is not accurate enough to choose the right OD pair and route. Hence in this paper we have used a semi-supervised route choice methodology to determine the accurate location of the vehicle by means of semi-supervised learning method. The sparsity issue is handled by using OD inference model and measurement equation in relationship with AVI data. Based on the simulations, we have identified that the proposed semi-supervised method is efficient in terms of accuracy and interpretability. Moreover, this proposed method also indicates the input data's robustness. Future work can be on identifying and examining the variability of the existing work with respect to different population group, different trip motives, different time and different days.

References

- [1] Mak, C. L. (2017, March). Passive UHF RFID tag designs for automatic vehicle identification. In *2017 IEEE International Conference on Computational Electromagnetics (ICCEM)* (pp. 61-63). IEEE.
- [2] Cui, Y., Xu, H., Wu, J., Sun, Y., & Zhao, J. (2019). Automatic vehicle tracking with roadside LiDAR data for the connected-vehicles system. *IEEE Intelligent Systems*, *34*(3), 44-51.
- [3] Ariff, F. N. M., Nasir, A. S. A., Jaafar, H., & Zulkifli, A. N. (2021). Comparability of edge detection techniques for automatic vehicle license plate detection and recognition. In *Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019* (pp. 891-910). Springer, Singapore.
- [4] Lin, N. H., Aung, Y. L., & Khaing, W. K. (2018, November). Automatic vehicle license plate recognition system for smart transportation. In *2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)* (pp. 97-103). IEEE.
- [5] Sridevi, T., Swapna, P., & Harinath, K. (2017, January). Vehicle identification based on the model. In *2017 IEEE 7th International Advance Computing Conference (IACC)* (pp. 566-571). IEEE.
- [6] Shirley, D. R. A., Ranjani, K., Arunachalam, G., & Janeera, D. A. (2021). Automatic Distributed Gardening System Using Object Recognition and Visual Servoing. In *Inventive Communication and Computational Technologies* (pp. 359-369). Springer, Singapore.
- [7] Al-bakry, A. M., Al-mamory, S. O., & Mushatet, H. H. (2017, March). Comparative study on automatic vehicle identification techniques. In *2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT)* (pp. 150-155). IEEE.
- [8] Shobayo, O., Olajube, A., Ohere, N., Odusami, M., & Okoyeigbo, O. (2020). Development of Smart Plate Number Recognition System for Fast Cars with Web Application. *Applied Computational Intelligence and Soft Computing*, 2020.

- [9] Manoharan, S. (2019). An improved safety algorithm for artificial intelligence enabled processors in self driving cars. *Journal of Artificial Intelligence*, 1(02), 95-104.
- [10] Bhalaji, N. (2019). Performance evaluation of flying wireless network with Vanet routing protocol. *Journal of ISMAC*, 1(01), 56-71.
- [11] Ahmed, M. M., & Abdel-Aty, M. (2013). Application of stochastic gradient boosting technique to enhance reliability of real-time risk assessment: use of automatic vehicle identification and remote traffic microwave sensor data. *Transportation research record*, 2386(1), 26-34.
- [12] Nantes, A., Ngoduy, D., Miska, M., & Chung, E. (2015). Probabilistic travel time progression and its application to automatic vehicle identification data. *Transportation Research Part B: Methodological*, 81, 131-145.