Estimating Vehicular Traffic Count Using Mobile Phone Network for Effective Traffic Management

By

E.N. Onwuka, M. David and *A. Abdulrasaq
Department of Telecommunication Engineering,
School of Engineering and Engineering Technology,
Federal University of Technology, Minna, Nigeria.
*Email: *amraz4real@gmail.com,

ABSTRACT

Obtaining quality information about traffic congestion for effective traffic management of a location requires an expensive means with tedious operations. Generally, existing methods such as the use of manual counting, global positioning system (GPS), radar, inductive loops are all costly with respect to time, personnel and complexity. In this paper, mobile phone flow is being measured instead of measuring vehicle flow (traffic flow). We made use of a k-means clustering technique on a network-triggered data to estimate the vehicular traffic count of a highway. Results of this technique was evaluated with the official count and gives a mean percentage error of 12.96%. This technique proves to be reliable and cost-effective in estimating traffic count which could be used on a large scale.

Keywords: clustering, mobile phone flow, mobile network data, traffic congestion, traffic management, vehicle flow.

INTRODUCTION

Advanced transportation systems are key characteristic of modern-day society which affects the individual's lifestyle, societal structure, and offer people convenience with freedom. Nevertheless, it also constitutes a problem of frustrating congestion and delays, especially in most urban areas, where traffic demand increases steadily and the transportation infrastructure seizes to expand at the same pace. These facts bring about the so-called traffic congestion problem.

In other to solve this problem, an effective traffic management technique such as traffic count (ground truth) is being carried out. Ground truth records approximates the number of observed traffic on roads and these records are gotten from the use of manual count or sensor usage by the transport authorities which is used in traffic modelling and transportation studies. The existing automatic traffic counting techniques such as road sensors, cameras, GPS are highly expensive, therefore making countries with lower economy to subscribe to using manual counting technique i.e. assigning some personnel to record the number of cars passing through the road. With respect to the ubiquity of cell phones in the world today, the use of mobile phone data as a proxy for traffic modelling is being introduced. Basically, using the mobility features in a call detail records CDR, a clustering technique is developed to automatically determine the traffic density on highway that is covered by a base transceiver station (BTS). And finally, the official traffic count (ground truth) is evaluated and correlated with the estimated mobile phone based traffic count. This method provides a dependable technique which ensures an affordable traffic count measurement on a large scale.

RELATED WORKS

Studies in [17], [18], and [22] showcase the identification of traffic status (road congestion and traffic route classification) using cameras or GPS as a conventional method. Low cost image processing algorithm (LIPA) algorithm was used in [20] to separate the real image of a vehicle from the whole picture background obtained from a camera, this helps to filter the capturing and improves the accuracy of the technique. J. Lovell David made use of mobile phone data to measure travel speeds and travel times in [16], Johan et al also obtained traffic information using handover patterns from UMTS signals [14, 4]. Handoff patterns from mobile phone networks is showcased in [13] to detect the routes that people mostly take through a city. Investigations on origin-destination matrix calculations [15] made use of simulations of mobile phone data. A month estimation in Portugal reflects a tight relationship between a mobile phone attribute and urban traffic network which was proven by the correlation between the densities of a taxi with mobile phone call variation [8]. Six models were proposed in [24] to estimate traffic density using an anonymised CDRs (calling behaviour of mobile phone users). This work attempts to use a clustering technique to estimate vehicular traffic counts from mobile phone flows.

THE DATASET

Network-triggered dataset are information of network subscribers collected by telecommunication companies for billing and routing purposes. This is influenced by the network itself which also permits an automatic update model of information. This information is gotten from the base transceiver station (BTS) which has its longitude and latitude attributes covering an area (cell). Every time a phone is ON or when a call is made or received, a set of variables are saved including the cell phone numbers, date and time, coordinates of phones. For this research, a day (24 hours) network triggered record for the UK highway was used.

In addition, an official traffic count of the same highway provided by the highway agency in the UK was used for comparison purpose (http://data.gov.uk/dataset/dft-eng-srn-routes-journey-times). And this contains the average traffic flow record, which was taken on every 15-minute interval throughout the day with the use of several devices like cameras, pyro-electric and piezo-electric devices etc positioned on the road link.

METHODOLOGY

Considering the processed network-triggered data (i.e. CDR data having solely mobility features like coordinates, distance, time), the speed of each phone on the highway was determined using the expression in the equation below:

$$VP_{it}^{j} = \frac{\sqrt{[x_{it}^{j} - x_{i(t-1)}^{j}]^2 + [y_{it}^{j} - y_{i(t-1)}^{j}]^2}}{T}$$

Where VP_{it}^{j} = speed of phone i in vehicle j during time t, $x_{i(t-1)}^{j}, y_{i(t-1)}^{j}$ and x_{it}^{j}, y_{it}^{j} are the position coordinates of phone i at time t-1 and t, T is the observation period from time t-1 to time t. [23]

A 10 mph speed threshold was set for a K-nearest neighbour classifier to classify the phone speeds into a moving speed value or a non-moving speed value. Due to some abnormalities in human being's walking speed, the 10mph threshold value was considered a safe value because normal human walking speed is given as 5.6 mph [32]. This was done specifically to differentiate a phone in a moving vehicle from a phone held by pedestrians which further aided the clustering process.

$$moving = \{i, with\, s > 10mph\}$$

Where s is the speed of phone and i is the corresponding phone.

K-Means Clustering

K-means clustering is a partitioning technique which works on real observations and creates a single level of clusters which makes it often more suitable for large amounts of data. It treats each observation in a data as an object having a location in space. It finds a partition in which data within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. K-means uses an iterative algorithm that minimizes the sum of distances from each data to its cluster centroid or centre (i.e. the point to which the sum of distances from all data in that cluster is minimized), over all clusters. This algorithm moves data between clusters until the sum cannot be decreased further and the result gives a set of clusters that are as compact and well-separated as possible.

Phone Clustering

An unsupervised k-means clustering algorithm was used to observe the outcome of the classifier i.e. moving or non-moving phones, it then identified phones with close attributes and organised them into groups based on location, speed and time, such that each group represents a cluster and hence, the total number of cluster equals the total number of vehicle (traffic volume). Phones in the same cluster have a very similar coordinate (location), speed and time difference [24].

RESULTS AND DISCUSSION

With the use of the k-means clustering technique, the aggregate flow of phones was detected and then clustered into groups based on the closeness of their mobility features. This was done for every period of observation (15 mins interval) for a whole day (i.e. 24 hours). Each clustered phone group means that the phones are together and thus equals one vehicle. The total number of clustered mobile phone flow in a period represents the total vehicle count for that period. Thus, for a day (24 hours), there are 96 periods and estimating the total number of vehicle count for the day is the sum of all the vehicle count per period.

Fig 3.1 shows the error plot for each period i.e. Period 1-96. It is the difference between the actual count (ground truth) and predicted count (clustered phone flow).

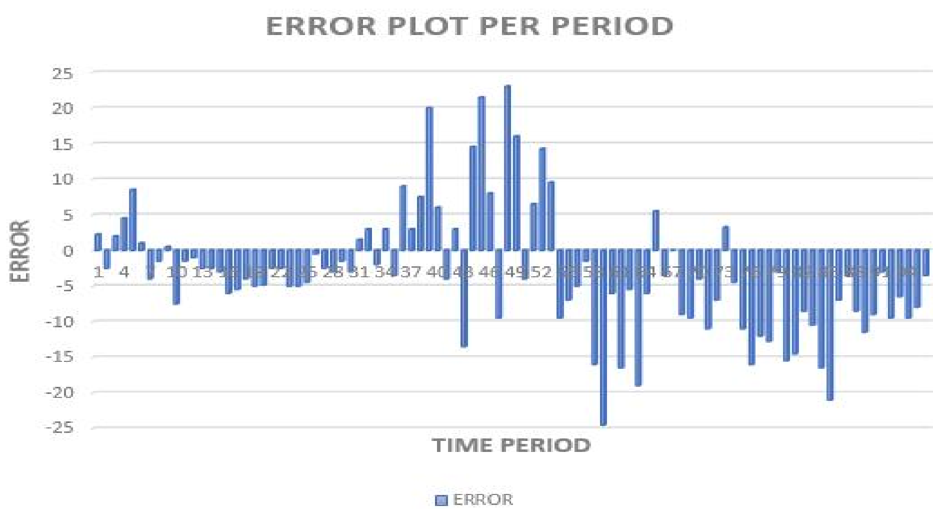

Fig 3.1: error plot per period

Fig 3.2 is a graph that shows the traffic trend of the highway for both actual vehicle count (ground truth) and the clustered phone flow for the whole day at every 15 mins interval.

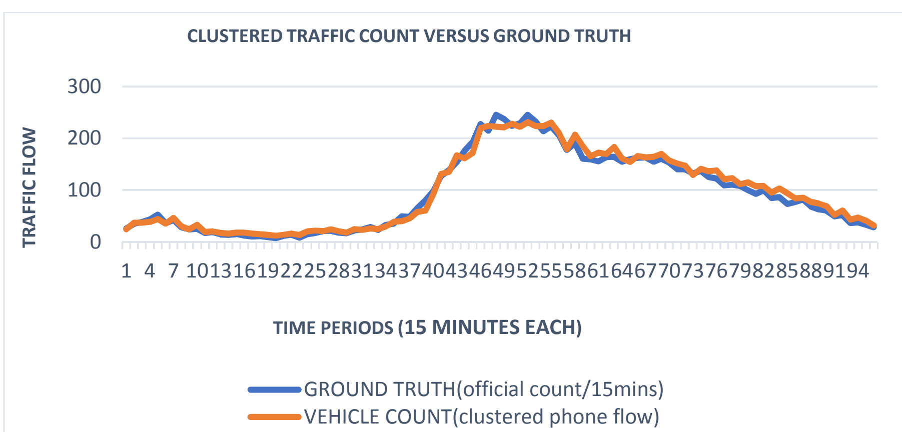

Fig 3.2: clustered vehicle count versus ground truth

The technique gave a mean percentage error of 12.963% and using Matlab 2009b, the correlation between the actual count and the predicted count gives $\boldsymbol{r} = \mathbf{0.99293}$ as shown in figure 3.3.

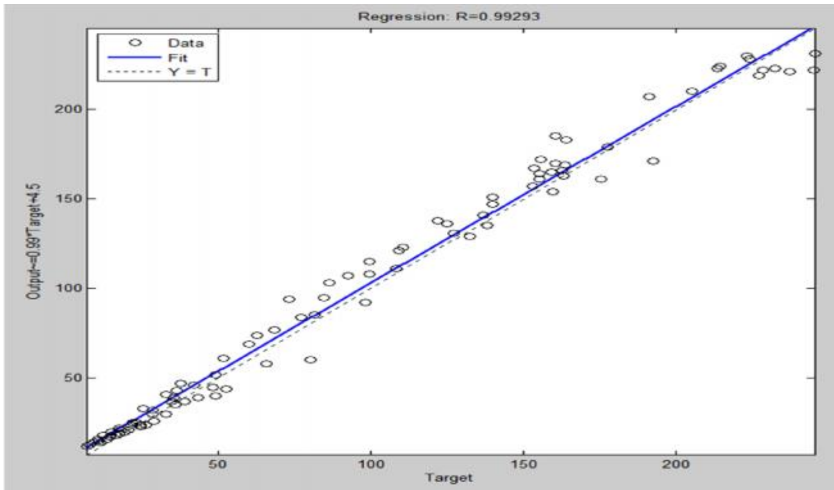

Fig 3.3: k-means correlation plot.

CONCLUSION

This study introduced clustering technique on phone user's mobility features in a network-triggered dataset. It considered a presence of more than one phone in a car by clustering phone users with close mobility features to enhance the detection of highway vehicular traffic count using mobile phone network. The results obtained indicates that considering the presence of more than one phone in vehicles proved to be more accurate and reliable rather than just observing phone flows as done by other techniques that made use of linear regression model or support vector regression [31], [32]. It is recommended that the collection of network triggered dataset should consider more mobility features like altitude of phones, as it would ascertain the suitability of estimating vehicular traffic count in an area or highway. Specifically, this paper is based on showcasing a simple technique with a recognisable degree of social impact which could estimate vehicular counts on highway with less resources.

REFERENCES

[1] D. F. G. A. Bolla R., "Estimating road traffic parameters from mobile communications," 2000.

[2] D. G. a. J. M. Karlson, Road traffic estimation using cellular network signalling in intelligent transportation system, Nova science publisher, 2009.

[3] D. B. M. S. a. M. M. florian Knorr, "Reducing traffic Jam via VANET," *IEE Transaction,* pp. 3490-3498, 2012.

[4] S. I. a. T. A. Najiba Ahmadullah, "Real-time Optimum Vehicle Routing using Phone Network," 2015.

[5] W. W. G. J. Peilan He, "Efficient Algorithms for Vehicle Type Identification Using Mobile-phone Locations," pp. 161-165, 2014.

[6] A. A. D. M. Ayalew Belay Habtie, "Cellular Network Based Real-Time Urban Road Traffic State Estimation Framework Using Neural Network Model Estimation," pp. 38-44, 2015.

[7] Y. S. T. L. Li Hao, "Research on The Evaluation of Traffic Condition Using Mobile Phones," 2010.

[8] S. P. a. C. B. Marco Veloso, "Exploring the Relationship between Mobile Phone Call Intensity and Taxi Volume in Urban Area," pp. 1020-1025, 2012.

[9] M. W. o. a. P. T. Chirayut Panmungmee, "Automatic Traffic Estimation System using Mobile Probe Vehicles," pp. 11-15, 2012.

[10] P. J. T. O. J. S. Markus Friedrich, "Mobile Phone Data for Telematic Applications," 2008.

[11] B. K. Mukta Ranjan Singha, "Mapping Mobile Phone Network onto Urban Traffic Network," in *International MultiConference of Engineers and Computer Scientists*, Hong kong, 2013.

[12] R. Horak, "Telecommunications and data communications handbook," *WWiley-Interscience,* p. 110–111, 2007.

[13] F. Calabrese, "Real-Time Urban Monitoring Using Cell Phones:A Case Study in Rome," pp. 141-151, 2011.

[14] N. C. F. G. Johan p. wideberg, deriving origin–destination data from cellular network, 2006.

[15] k. sohn, "Dynamic Estimation of Origin-Destination flows Using Mobile phones," 2004.

[16] J. Lovell David, "Accuracy of speed measurements from cellular phone vehicle," pp. 303-325, 2001.

[17] T. R. Currin, "Turning Movement Counts. In Introduction to Traffic Engineering: A Manual for Data Collection and Analysis," pp. 13-23, 2001.

[18] S. T. a. R. Rajamani, "Portable Roadside Sensors for Vehicle Counting, Classification, and Speed Measurement," pp. 73-83, 2014.

[19] M. N. A. C. J. B. O. O. a. P. S. Luis Unzueta, "Adaptive Multicue Background Subtraction for Robust Vehicle Counting and Classification," *IEEE TRASACTION,* pp. 527-540, 2012.

[20] L. Z. K. X. R. S. a. L. C. Rui Wang, "EasiSee: Real-Time Vehicle Classification and Counting via Low-Cost Collaborative Sensing," *IEEE TRANSACTION,* pp. 414-424, 2014.

[21] R. Z. a. X. Wang, "Counting Vehicles from Semantic Regions," *IEEE Transaction,* pp. 1016-1022, 2013.

[22] S. T. a. R. Rajamani, "Portable Roadside Sensors for Vehicle Counting, Classification, and Speed Measurement," *IEEE Transaction,* pp. 73-83, 2014.

[23] G. H. a. L. Fasheng, "Estimating freeway traffic measures from mobile phone location data," *European Journal of Operational Research,* pp. 252-260, 2013.

[24] L. M. R. F. G. B. a. J. M. d. C. Noelia Caceres, "Traffic Flow Estimation ModelsUsing Cellular Phone Data," *IEEE Transaction,* pp. 1430-1441, 2012.

[25] W. P.-a. a. R. Peachavanish, "Estimating Road Traffic Congestion from Cell Dwell Time using Neural Network," pp. 1-6, 2007.

[26] F. A. C.-P. a. H. -V. Andrés Rico-Páez, "Teletraffic Analysis Formulation based on Channel Holding Time Statistics," pp. 326-330, 2009.

[27] J. Wang, "Design of Mobile Phone Networks and Map-matching Navigation System," *Elsevier,* p. 2045 – 2049, 2012.

[28] V. A. a. G. Guido, "MOTORWAY TRAFFIC MANAGEMENT AND TRAFFIC PARAMETERS ESTIMATION FROM MOBILE PHONE COUNTS," pp. 534-538, 2002.

[29] K. K. a. H. Choi, "A Mobility Model and Performance Analysis in Wireless Cellular Network with General Distribution and Multi-Cell Model," p. 179–198, 2010.

[30] T. M. C. N. T. N. L. D. T. a. H. D. T. Hoang Van Dung, "A Vehicle Positioning and Navigation System Design Using Mini2440 ARM and Linux," *IEEE,* pp. 74-78, 2013.

[31] V. Frias-Martinez, (2014). Estimation of traffic flow using passive cell-phone data. In Workshop on Data Science and Macro-Modelling (DSMM), SIGMOD, 52-55.

[32] Tony, L. (2015). Cars and calls: using CDR data to approximate official traffic counts. 23-25.

