Estimation of traffic flow using passive cell-phone data

Vanessa Frias-Martinez
College of Information Studies
University of Maryland,
College Park
vfrias@umd.edu

Younes Moumni
EPFL
Laussane, Switzerland
benyounes.moumni@epfl.ch

Enrique Frias-Martinez
Telefonica Research
Madrid, Spain
efm@tid.es

ABSTRACT

In this paper we present preliminary results for estimating traffic flow using passive cell-phone network information. Two datasets are considered: (a) passive cell-phone data and (b) information provided by the English Highways Agency. Our proposed method identifies cell phone users that are traveling by car and using a linear regression model, estimates the flow for each of the links in which the road network is divided. Initial results indicate, that, under certain conditions, traffic flow can be effectively approximated with passive network data.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

1. INTRODUCTION

Cell phones have become pervasive sensors of human behavior. As a result, they can be used to characterize a variety of human-related activities, such as mobility or traffic. Although there are specific pervasive networks for monitoring traffic *e.g.,* CCTV networks, these are limited to just a few critical parts of the whole road network. However, cell-phone infrastructures, which are deployed for communications, typically cover much larger areas. Such network infrastructures are continuously collecting information about the location of cell-phones in order to be able to route calls when they happen. That information -not triggered by user requests- is what we call passive information of the network. The location is specified as the longitude and latitude of the cellular antenna that collects the information.

Traffic estimations can be done with a variety of techniques and sensors, such as cameras, toll information, road sensors, etc. each one suffering from a number of limitations. The literature has already focused on using cell-phone data for estimating traffic [1]. Although the best part of those studies use active information (CDRs or handovers), there are also studies that focus on using passive information [2] [3]. Nevertheless, such approaches focus on identifying road congestion in order to raise traffic alerts.

The goal of this paper is to evaluate whether we can use passive information from a cellular network to estimate traffic flow (as defined by Road Agencies). This opens the door to using our method for estimating road conditions in areas where a cell phone network is deployed but no infrastructure to estimate traffic has or can be deployed.

DSMM'14, Jun 27-June 27, Snowbird, USA
ACM 978-1-4503-3012-1/14/06.
http://dx.doi.org/10.1145/2630729.2630739.

2. DATA SETS

Passive or Network-driven cell-phone data captures periodic location information of network users. Triggered by the network itself, it allows to have an updated model of the network load. The location is captured in the form of a BTS (Base Transceiver Station) antenna which has a longitude and latitude associated *i.e.,* no information about the location of a user within the area of coverage of an antenna is known. The coverage of a BTS, called a cell, is approximated using Voronoi tesselation. For this paper, we use one day of encrypted passive network data for the UK.

Additionally, we use flow information provided by the Highways Agency in the UK as ground truth. Specifically, we use the journey time and traffic flow data series which contain traffic flow information every 15-minute interval on all motorways and 'A' roads managed by the Highways Agency (further information at http://data.gov.uk/dataset/dft-eng-srn-routes-journey-times). The time series are provided for each link (and direction) in a road, where a link is defined as the road between two junctions. Flow is estimated using a combination of sources, including Automatic Number Plate Recognition (ANPR) cameras, in-vehicle Global Positioning Systems (GPS) and inductive loops built into the road surface. The flow value is estimated every 15-minutes and computed as the average number of cars per minute at a given link during that time interval.

3. TRAFFIC FLOW ESTIMATION

The previous two datasets have different granularities, while the cell-phone data considers cell phone towers (and their coverage), road information is presented as a time series for each link in the road network. To approximate flow information from passive cell-phone data, we first need to merge the *geographies* of the two data sources. Figure 1 presents an example of how a link usually expands over a number of tower coverage areas (cells). Thus, we first associate BTS towers (cells) to road links by identifying the areas of coverage that intersect with each link. For each set of cells covering a given link, we compute the convex hull and refer to it as the envelope of the link.

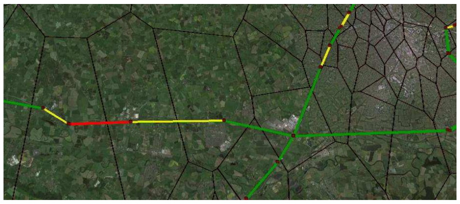

Figure 1: Example of overlap between the cells of the cell-phone network and the links.

To approximate the traffic flow of a link during a specific time interval, we count the number of phones that might be traveling by car through the envelope of the link in that period of time. For that purpose, we compute for each phone in our dataset: (1) the pairs (A, B) of cell transitions; (2) the time (t_A, t_B) when the phones were at those cells. However, in order to identify whether a given phone represents a car trip rather than a person walking, we require the speed to be within a range [*minSpeed*,*maxSpeed*]; where the speed is computed as the quotient between the distance from A to B and the time taken to travel such distance. Additionally, we also need to model the direction of the flow from the cell-phone data. Figure 2 presents an example where the envelope of the link is crossed by a cell transition (A, B). Following our method, the cell transition would count towards the flow of the link in the envelope. However, it is not clear in which direction of the traffic. To elucidate direction, we define a $maxAngle$ parameter that determines the maximum angle between the cell transition line and the link to be considered flow in a given direction. We empirically determine the values for the parameters in our method, by looking for the best fit between our model and the ground truth.

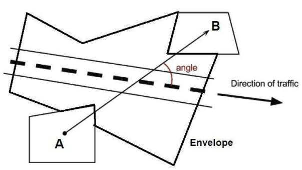

Figure 2: Example of (A,B) segment crossing the envelope of a link and the angle defined by the direction of the traffic.

Figure 3(left) presents the best fit: the x-axis shows the count per link using our flow approximation method ($fcount$) and the y-axis the actual number of cars in that link. The linear regression has an $R^2 = 0.35$ with $RMSE = 157$. We observe that our method seems to overestimate or underestimate the actual number in several occasions, which was expected due to the nature of the data. More specifically, the largest estimation errors appear when the real number of counts is large, which we underestimate. An in-depth look into the types of links reveals that our largest estimation errors happen in motorways, whereas the smaller underestimations are for roads. Thus, to take advantage of the strength of our model, we exclusively focus on flow analysis for roads. Figure 3(right) presents the linear regression considering only the road links, and we can see how in this case the number of links that are under or overestimated is reduced. The fit has an $R^2 = 0.65$ and an $RMSE = 70$, a considerable improvement from the previous case. Also, it can be observed how, as expected, road links have less flow that motorway links. The method parameters for the best fit were $minSpeed = 20km/h$, $maxSpeed = 200km/h$ and $maxAngle = 45deg$.

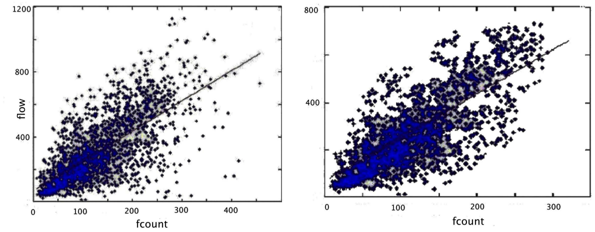

Figure 3: The x-axis presents the count per link ($fcount$) and the y-axis the actual flow (left) for all the links and (right) for the roads links; with the corresponding linear fit.

Using the regression model, we compute the flow for ten road links in the UK. Figure 4 presents the real and the estimated flow for each link during a period of 24 hours (in 15-minute intervals). The x-axis shows the temporal series for each link and the y-axis the value of the flow (as cars per minute). The average normalized difference between the original and estimated flows across the ten links is 0.83. These results seem to indicate that flow can be effectively estimated with passive information, at least for road links.

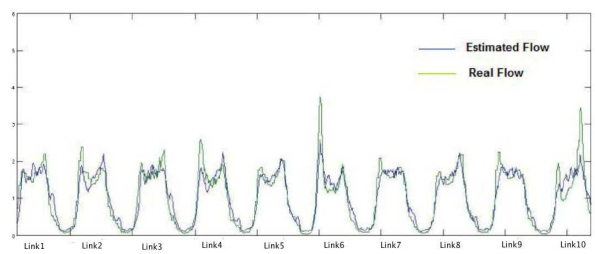

Figure 4: Temporal flow series for each one of the 10 links considered: original and estimated.

4. REFERENCES

[1] C.-Y. Chiang and et al. Estimating instant traffic information by identifying handover patterns of umts signals. In *Intelligence Transportation Systems, 14th Int. IEEE Conf.*, pages 390–395, 2011.

[2] A. Janecek and et al. 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*, pages 361–370. ACM, 2012.

[3] D. Valerio and et al. Road traffic estimation from cellular network monitoring: a hands-on investigation. In *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pages 3035–3039. IEEE, 2009.