The development of intermodal transport services on the basis of geoanalytical information

Denis Vlasov¹*, Anton Menshutin¹, and Nina Danilina²

¹Genplan Institute of Moscow, 125047, 2/14, 2-nd Brestskaya str., Moscow, Russia
²National Research University Moscow State University of Civil Engineering, 129337, 26, Yaroslavskoye Shosse, Moscow, Russia

Abstract. The article highlights the application of geoanalytical data for the purposes of urban transportation planning. It presents the original method for functional analysis and development of public transport transit hubs based on the study of cellular network information obtained from mobile operators, as a foundation for development of the effective intermodal system for public transportation services. The approbation of the suggested methodology was carried out using the results of analytical research of the data obtained from mobile network operators, serving the territory of Moscow. The results of research refer to one of the stations of the Moscow metro. Studying motion information of residents and workers provided by signal transmission from mobile communication network allows the authors to describe the area of major importance for the transit hub, needed to make valid design solutions at the stages of planning the effective work of the urban transport system.

1 Introduction

Effective management of a large metropolis represents an enormous challenge for the city government. The vastly diverse city assets are interconnected in most intricate ways. Changes applied to one part of the system can produce complicated results, having an effect on the entire city. In order to manage system that complex, a vast amount of data is required. Since a lot of such data essentially describes human behaviour, and because measuring human behaviour on a grand scale is very difficult, if not altogether impossible (at least within the realm of ethical and legal norms, and budgetary restrictions), an accessible proxy to direct behavioural metrics is often desired.

The sphere of public transportation services is among those most important city assets that do need such kind of information. The formation and development of convenient and safe system for intermodal transportation is an essential factor for sustained development of any settlement, especially the largest cities, agglomerations, conurbations and other contemporary forms of settlement. Intermodal transportation is understood by the authors as a well-developed system of passenger transportation, thus offering best possible

* Corresponding author: vlasych@mail.ru
coordination of different modes and comfortable transfer [1,2,3]. So, the main objective of this paper is to examine the possibility of implementation of geoanalytic data in urban planning studies. In particular, it is determining the prospects of intermodal transport infrastructure prosperity and the subsequent progress of "big data" processing methods on the simulation of the transport system development.

2 Methods

The critical issue for determining the prospects for the development of intermodal transport systems is that of valuation of traveler volumes along transit network, incoming and outgoing passengers for each transit hub. The affected area of a Transport Transit Hub (TTH) is a residential and/or employment area gravitating to it.

To date national urban development practices several analytical methods for calculating the areas affected by the Transport Transit Hub [4,5]:

- regulatory methods based on requirements on time or distance approach to TTH;
- graphic-analytical method of «isochrones» definition based on the speed analysis of major highways that ensure delivery of population to TTH;
- statistical method based on the analysis of load level of the ground public transport routes that ensure the delivery of the population to TTH.

The main disadvantages of the above methods are either small precision (the first and the second methods), or significant labor input (the third one). Therefore, there is a real need for the use of modern technologies for determining the area affected by one or another TTH [6,7]. Geoanalytical study of the obtainable data from cellular networks operators represents one of them.

Mobile phone operators collect a lot of technical data about operation of Base Transceiver Stations (BTS's) [8,9]. By analyzing this data, one can estimate, among other measurements, the total number of mobile devices at a given territory at different time spans, the number of trips between a pair of regions.

In this paper we show how mobile network data can serve as such a proxy. In particular, we describe how it can be used for the analysis and development of Moscow public transport transit hubs.

City occupies large portion of space and its infrastructure is very irregular in space. Thus, data that is required to describe the city should be spatial. Cellular networks collect device location data in order to provide service. Precision of this device location data varies with time and space. This data is collected at irregular time intervals. Events that trigger device location inquiry are either device specific or network specific. Device specific events are incoming or outgoing voice calls or SMS’s, and device being switched on or off. Network specific events occur due to network related activity – crossing of Location area boundary or Periodic location update after extended period of inactivity. The frequency at which device location data is gathered is subject to exogenous factors and can’t be influenced under normal course of operation since only passive collection is possible.

3 Implementation

In our study we use data on mobile network events in Moscow and Moscow region covering a period of one month. Total number of mobile devices covered by this study is approximately 20 million. For each event its date and time, cell and lac id, and device id were known. Coverage area for each cell was estimated, defined as locus of stable device-
to-cell linkage. Coverage area size may vary from few hundred meters to tens of kilometers, depending on the topography, building density, number of mobile subscribers within that area and types of devices they use, and some other factors. During the event the device location is recorded as particular cell coverage area.

In order to comply with privacy regulations, all data processing was implemented on cellular operator side. Data collection, algorithm implementation, and final processing were all carried by operator’s IT team, thus preventing access to private data by the third party.

Data processing proceeded in three stages: pre-filtering, main processing, and data aggregation. Pre-filtering stage is straightforward: data from devices other that mobile phones (e.g. ATM, tablets, navigators, etc.) is removed. During the main processing stage, data is transformed from event-specific to location-specific. Since we are interested in the actual locations of devices and not in network events per se, chains of consecutive data entries generated by one device pointing to the same location were aggregated. Apart from obvious chains of events occurring within the same cell area, slightly more sophisticated filtering techniques were used, designed to detect situations where an unmoving mobile device registers cell change. This occurs e.g. when a device located near a border of two adjacent (or overlapping) coverage areas switches back and forth between those cells. This produces consecutive events where cell changes from, say cell #id1 to cell #id2 back and forth several times over a short period of time. If coverage areas of those cells overlap, then we assume that device is at a stop within that overlap area.

We next estimate place of residence and place of employment for each subscriber. Place of residence is defined as an area where subscriber spent most of the time during night hours through the month. Similarly, place of employment is defined as a location where subscriber spent most of the time during the working hours, and that is different from his place of residence. If those two locations match, than subscriber is assumed to be unemployed.

A mobile device that remained within the same area for 60 minutes or longer is assumed to be stationary for that period of time; we call that occurrence a static event. Device relocation occurs when the device moves from one cell area to another. A trip is defined as a consecutive sequence of device relocations, preceded and followed by two static events.

Trips that were made on metro were categorized into separate group. Technically it was possible as most of the metro network is underground, and id’s of cells serving metro stations and tunnels are known. Therefore, with high probability, a subscriber entering the metro will register with a metro-serving cell, and will be dropped upon exiting metro. More complicated algorithms were utilized to account for above ground parts of metro, especially to identify trips that started or terminated at such stations.

From analytical perspective, separating metro trips from others is useful for two reasons. First, metro is the main mass-transit system in Moscow, the number of trips made on metro is high – almost 9 million during any given business day. Second, it allowed for additional verification of our transit estimates with metro-transit data.

Transformation of data to location-specific form and identification of trip-related events allows researcher to estimate variety of parameters that closely approximate behavior of people within the city, such as number of inhabitants residing within given region, number of trips between a given pair of regions, and so on.

4 Experiments

Data-aggregation stage served two purposes: one was to reduce the amount of data, the other was to preserve privacy. The choice of aggregation schemes depended on the
specific topic of the report. Several reports on various topics were produced. Here are few examples of aggregation techniques. Let assume that place of employment and place of residence are known for each device. If we aggregate the database over the place of residence we can estimate the population of the given region. If we aggregate data over two columns we can estimate commute matrix.

Transportation hubs analysis required aggregation of data to square grid of 500 by 500 meters cells. If a cell coverage area overlapped several grid cells, we randomly assigned its data to one of those grid cells, with probabilities proportional to intersection areas. In pre-aggregated form, data contained the following attributes for each trip made using Moscow metro:

1) Origin and destination metro stations.
2) Grid cell at which place of residence is located (if exists).
3) Grid cell at which place of employment is located (if exists).
4) Origin and destination grid cells.

The approbation of the suggested methodology was carried out for example at one of the stations of the Moscow metro at Tagansko-Krasnopresnenskaya line – station Polezhaevskaya, located at Moscow Northern administrative district. It is one of the busiest transport terminals enabling the number of linkages between land and off street transportation modes. Customers’ access is provided by 7 bus and 6 trolleybus routes, as well as by private car and by foot. Objective of the research was to highlight the affected area of the TTH on the basis of the data of mobile operators on the use of station by residents and employees.

The results of the research at Polezhaevskaya Metro station are presented in Figure. 1, a,b.

The final aggregated report shows following attributes for each metro station:
1) Number of people residing within each grid cell.
2) Number of people being employed within each grid cell.
3) Number trips started from within each grid cell.
4) Number trips terminated at within each grid cell.

Frequency for all data is hourly. All estimates were done from 8 AM to 10 AM in work days, April 2016.

Our estimates show fairly good qualitative agreement when compared with results from population databases and transportation surveys: our estimates closely replicated dynamics of the real world data, namely periods of highest and lowest traffic loads (in both, metro and above ground transport). The quantitative agreement, however, wasn’t as good: estimate error was usually within 10%-15% range. We attribute those discrepancies primarily to inaccuracies in location estimation.

5 Results

The research results reported in this article show that using a geoanalitical data offers the possibility to define the area affected by TTH and not only. It provides also unprecedented opportunities for analysis and studies in the other fields of urban transportation, such as:
- further justification for the classification provisions of transfer hubs to determine the priority directions for development of city transport system;
- TTH load forecast (which requires further research and identifying settlement patterns);
- various applied works, related to routings and frequencies of land public transport services basing, frequencies on off street high-speed lines and so on and so forth.

Development of methods related to data analysis of cellular operators, will significantly improve the accuracy of urban calculations and forecasts, provide municipalities with operational information on places of settlement and employment.

References