Events that Affect Urban Mobility Patterns: an Analysis of Beijing GPS Data

Edson B dos Santos Jr
Department of Informatics
Universidade Federal do Estado do Rio de Janeiro
Rio de Janeiro, Brazil
edson.santosjr@uniriotec.br

Carlos Simões
Department of Informatics
Universidade Federal do Estado do Rio de Janeiro
Rio de Janeiro, Brazil
carlos.simoes@uniriotec.br

Ana Cristina Bicharra Garcia
Department of Informatics
Universidade Federal do Estado do Rio de Janeiro
Rio de Janeiro, Brazil
cristina.bicharra@uniriotec.br

Adriana S Vivacqua
Department of Computer Science
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brazil
avivacqua@dcc.ufrj.br

***Abstract*— The analysis of data collected through GPS and / or smartphone logs allows the identification of human mobility patterns. This identification has been the subject of several surveys in recent years. However, certain factors can change these patterns, such as special events of great magnitude (e.g., the Olympic Games) or environmental events (e.g., an increase in air pollutants.) In this paper, we present an investigation into the impact of these events on these patterns. We analyze urban mobility data collected in Beijing between April 2009 and October 2012, we use Artificial Intelligence techniques to identify mobility patterns in these data and then check for changes in patterns during the aforementioned events. The analysis of these data has become an important tool to extract insights and to support the planning of the urban displacement in the day to day and in special events.**

Keywords*— spatial data mining, GPS trajectories, air pollution, urban mobility, events that affect mobility patterns.

I. INTRODUCTION

Urban mobility is the condition in which the movements of people and cargo take place in the urban space of a city, urban agglomeration and/or metropolis. It equates to ease of moving within the city. Smart cities make use of large amounts of data from various sources in order to identify and understand human mobility patterns and socioeconomic phenomena [9]. Adequate transportation and circulation policies to improve urban mobility are related to continuous planning and ecologically sustainable infrastructures, and seek to produce effective and proportionate benefits to the resources used.

Our literature review identified that several authors considered the issue of mobility data collection a factor of great importance. In the past, metropolitan planning agencies needed a large amount of resources to conduct research in order to identify how and where individuals move. Data collection was done in small samples (about 1%) and low frequency (once or twice in a decade) [7], making it difficult to identify movement patterns.

In the last decade, many techniques have been proposed to collect, process, manage and mine displacement trajectory data, promoting a wide diversification of applications in distinct and related areas [5], which make use of data associated with different modes of transportation. These describe human movements with different purposes and at different spatial scales. Understanding such patterns is significant for urban mobility modeling and design of more efficient means of transport, serving the largest possible share of the population [8].

Changes in the way people interact with the Web and the increasing availability of GPS-enabled devices lead to a large number of GPS trajectories, representing people's location histories [1] [2] and enabling the creation of models that seek to identify patterns. These patterns are influenced by events of great magnitude, such as large scale sporting events or environmental events, such as increasing the amount of pollutants in the air. Analyzing large volumes of data over time, measuring the effects of unplanned events such as demonstrations, accidents, weather phenomena, and failures in public transport can provide input to support urban planning and to create a city that is inviting to people, and at the same time, conducive to sustainable urban mobility.

Individuals have specific urban mobility routines. Individual changes in these routines are expected to accommodate specific needs. However, collective routine patterns' changes may show the crowd's reaction to some special circumstance. A change provides more information whenever the aggregate routing changes converge to specific locations and side effects might occur, such as, increased pollution rates or traffic congestion. Monitoring the crowd to detect changes can reduce adverse side effects by enabling a quick response. In this sense, the use of Artificial Intelligence has promising potential by allowing the identification of models and patterns of urban mobility.

The next section presents a brief literature review. Section 3 presents the Artificial Intelligence and Machine Learning techniques used to identify patterns of urban mobility and the impact of events of great magnitude such as the XXIX Olympic Games, Beijing, 2008 and a record pollution index recorded in 2009. Section 4 presents the results of applying the techniques described in section 3, illustrated using maps. Section 5 presents the conclusions.

II. LITERATURE REVIEW

This section presents work related to the identification, collection and analysis of urban mobility data. In this review we identified that several factors that favor the collection and analysis of mobility data and factors that hamper such activities. Jiang, Ferreira, and Gonzalez [7] state that metropolitan planning agencies in the past required a large amount of resources to conduct research, with the aim of identifying where and how individuals move. The data collection was done considering small samples (something like 1%) and low frequency (once or twice in a decade).

Understanding human mobility patterns has gained extraordinary attention because of the rapid development of location acquisition technologies, complex network sciences, and human dynamics [8] [11] [16]. Mobile computing and GPS enable people to move and interact with information, services, and other people [10]. The analysis of mobile phone GPS data enables the extraction of human mobility patterns for millions of anonymous residents in a metropolitan area and generates knowledge through the analysis, filtering, scaling and understanding of different patterns of human mobility in cities and neighborhoods [7] [16]. With the enhancement of storage technology and big data processing, mobile operators can extract and store a large amount of behavioral data and develop smart applications.

Traffic sensing uses human mobility patterns learning techniques from updated location information in network Understanding mobility patterns benefits the design of transport policies [7]. Data mining techniques provide a way to make the knowledge of the trajectories available to researchers and professionals from the different communities that deal with urban mobility [6].

There are models and patterns that have been defined with the objective of understanding the regularity and variability of urban mobility based on the analysis of data collected by GPS. Zheng, Y., Li, Q., Chen, Y., Xie, X., Ma, W. [8], used transport modes from GPS records based on supervised learning. They identified and included a set of features such as rate of change of course, rate of change of speed, stop rate, speed and simple acceleration referring to the movements of users. They also proposed a graph-based post-processing algorithm to improve inference performance.

Liang, Y., Zhou, X., Guo, B. and Yu, Z. [11], explore the relationship between temporal factors, work and age with mobility patterns. To detect the regularity and variability of the geo-trajectories, they proposed a model to extract the spatial interaction matrix, which provides the detection of socio-geographical boundaries, based on the cluster displacement algorithm, the interaction force and semantics to indicate the interaction between crucial spatial regions.

The model proposed by Zhao, X., Xu, T., Fu, Y., Chen, E. and Guo, H. [5] considers a Spatio-Temporal real-time multitasking regression learning structure (stfMTR) for best inference air quality index (AQI), intra-station time dependencies and inter-station spatial relationship. A new metric is incorporated to correlate the stations considering geographic distance and profile similarities. Extensive experiments validated the efficacy of the structure that prove the space-time potential in AQI inference.

Zheng. Y., Zhang. L., Xie, X., Ma, W. [1] used GPS trajectories generated by several users, extracting interesting sites, experience and classic travel sequences within a specific geospatial region, to understand the correlation between users and sites and enable the recommendation of travel and tourist guidance by the mobile device.

The analysis by Kang, C., Sobolevsky, S., Liu, Y. and Ratti, C. [8] compares human mobility patterns based on two distinct data sets and proposes an integrated approach to the use of taxis and mobile phones for more insights into population dynamics, transport, and urban settings.

Xia, F., Wang, J., Kong, X., Wang, Z., Li, J. and Liu, C. [9] propose an integrated analysis method to find human mobility characteristics in Shanghai, China. The approach uses methods to adjust human movement patterns based on subway traffic data and taxi GPS data. The proposed method can also be used for other traffic trajectory data, such as cellular calls, social media check-ins, and private car data. Through qualitative and quantitative analysis, they noticed that the displacement of the trip is more adequate to the log-normal distribution than to an exponential model.

Kim, M., Kotz, D. and Kim, S. [10], used mobility features to develop a software model that generates realistic mobility ranges for the user, comment that pause, direction and speed distributions of the movements follow a log-normal distribution. Zheng at al. [2] propose an approach based on supervised learning to infer people's movement modes from their GPS records. In addition to proposing a graph-based post-processing algorithm to further improve inference performance, the algorithm considers both the common real world constraint and the typical behavior of the user based on location in a probabilistic way. Using GPS records collected by 65 people over a 10-month period, he evaluates the approach through a set of experiments. As a result, based on the point-of-change-based segmentation method and decision tree-based inference model, the new features brought an eight percent improvement in inference accuracy over the previous result, and post-processing based chart achieved a further 4% improvement.

Huang and Xiao [14] analyze the potential challenges and opportunities in intelligent traffic detection from a data science point of view with data generated by mobile networks. First, it classifies the data resources available in the commercial radio network according to different taxonomic criteria. Describes problems in detecting traffic based on the mobile user's network log data. It then studies algorithms for processing and learning data in the extraction of traffic condition information from a large amount of mobile network registration data.

The analysis of urban mobility data supports planning of collective transportation to reduce the use of individual transportation [4]. Cars, motorcycles and trucks are some of the major sources of pollutant emission in the atmosphere. Removing them from the roads has a direct consequence in the air quality indices, due to the increase of carbon dioxide and residues of the fossil fuels in the atmosphere.

The pollution problem worsens when special events occur in the area, causing traffic congestion and leading to environmental degradation [7]. Urban air quality varies non-linearly and depends on multiple factors, such as meteorology, traffic, land use and urban structures [5]. Information on real-time air quality, urban mobility and the dynamics of the city are of great importance to support the control of air pollution and protect human health [4] [5].

III. METHODOLOGY

With the objective of understanding urban mobility patterns and changes in these as a function of special events, the following set of data were analyzed:

- Urban Mobility: GPS trajectory data as described in the User Guide Version 1.3 [1] [3].
- Air Pollution: Air pollution data were collected from the World Air Quality Index project.

The following are presented as the data were described, formatted, treated and analyzed, including the challenges faced and the experiments carried out in the search to understand the urban movement, the collective behavior in special events and the change of the urban mobility pattern caused by the convergence to specific locations.

A. Data Description - Urban Mobility

GPS trajectory data was obtained from the Geolife project (Microsoft Research Asia). The dataset contains data from 182 users, over a period of more than five years (April 2007 to August 2012). A GPS trajectory data is represented as a sequence of date and time points, each containing latitude, longitude and altitude information. The dataset contains 17,621 trajectories with a total distance of 1,292,951 km and a total duration of 50,176 hours. These trajectories have been recorded by different GPS recorders and GPS phones and have a variety of sampling rates. 91.5 percent of the trajectories are recorded in a dense representation, and every 1-5 second or every 5-10 meters per point [1] [2] [3].

The data set recorded a wide range of outdoor movements of users, including not only life routines like going home and going to work, but also some entertainments and sports activities such as shopping, sightseeing, dining, hiking, and biking. Although this dataset is widely distributed in more than 30 cities in China and even in some cities located in the US and Europe, most of the data was created in Beijing, China.

The dataset has been organized into folders, each storing a user's GPS log files that have been converted to PLT format. Each PLT file contains a unique path and is named by its start time. To avoid possible time zone confusion, we use GMT on the date / time in the property of each point, which is different from our previous version.

B. Data Description - Air Pollution

Air pollution occurs when dangerous or excessive quantities of substances, including gases, particles and biological molecules, are introduced into the Earth's atmosphere. Air pollution indoors and low urban air quality are listed as two of the worst toxic pollution problems in the world. It can cause diseases, allergies and also the death of human beings; can also cause damage to other living organisms, such as animals and food crops, and can damage the natural or built environment.

The air pollution index data used during the production of this article comes from historical records of the World Air Quality project (aqicn.org). These data are organized in a CSV file where the measured values of µg / m3 of airborne particles (PM 2.5 index) are recorded every hour of each day of a given year. PM 2.5 rates the amount of 2.5 micron diameter particles that, when inhaled, settle in deep lung sites such as alveoli and bronchioles.

C. Challenges in Data Handling

A major challenge was the preparation of the data for the analysis due to the format and initial structure of the data. As previously mentioned, the trajectory data originated from the latitude and longitude record, as well as the date and time, of the location of 182 individuals performed between April 2007 and August 2012, at each interval of 1 to 5 seconds. Despite this range in the collection interval, the GPS were not connected during the whole time of the collection, being at the discretion of each individual its activation. Figure 1 shows the distribution of individuals who collected some GPS data on each day of the experiment.

The preparation of the mobility data consisted of an ETL process, where the dataset of each of the 18,739 files was stored in a SpatiaLite database. SpatiaLite was chosen for its geo-referenced representation of the data, allowing its use with Geographic Information Systems. Additionally, it is open source and allows queries using the SQL language.

The air quality data, as previously mentioned, were recorded by the World Air Quality Index project and stored in CSV files with indication of the date and time of each quality measurement. Pollution data refers to the city as a whole. That is, a single index for the city as a whole at a particular date and time. These data were also stored in the same mobility data database.

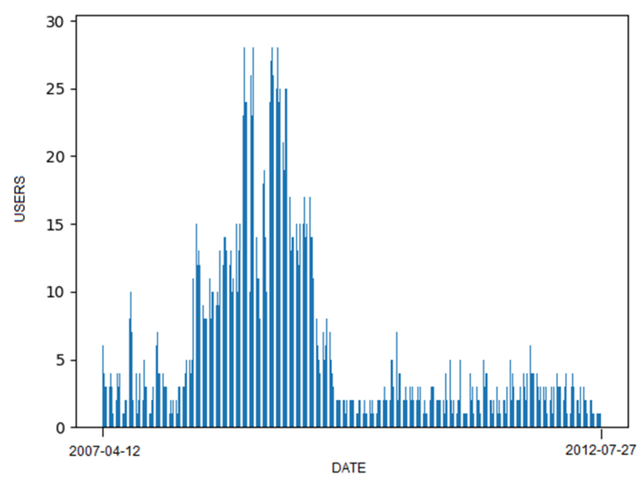

Figure 1 - Distribution of individuals who collected GPS data

D. Identification of Routine Displacements

To identify changes in patterns, it was necessary to identify patterns of routine displacements. Once the data were stored, the Machine Learning algorithm was used to identify frequent individual trajectories, as follows:

- Once the period of interest (Olympic Games or Pollution Peak) was determined, we identified the set of individuals who had data recorded in this period.
- Then, we enlarged the period of interest including three months before and three months after.

- From this data, we removed the date and time information, keeping only the coordinates, but considering the sequence.
- On the resulting dataset, we applied the k-means (unsupervised machine learning algorithm) to cluster the data points. As a result, several coordinate clusters were obtained.
- Then, we selected the clusters with the largest number of coordinates, and enriched with date and time information for each coordinate, for obtaining the daily routes.
- We made a new grouping of the daily routes according to the number of points in common, where we identified that the routes that had more than 70% of the points recorded in common also had the times of collection of these points very approximate, with a maximum difference of 19 minutes and a minimum of 8 seconds. We assumed, for our research, these routes are the frequent trajectories of each individual. It is interesting to note that all identified frequent trajectories were performed on weekdays rather than on weekends.

In an attempt to improve the identification of routine displacements, Latent Class Growth Modeling (LCGM) was used, which according to Apparicio, Riva and Séguin [13] presented better performance in the clustering of trajectories.

The same methodology used previously with k-means was used, this time finding an average of 76% points in common between distinct displacements in the clusters with the highest number of points. The same distinct displacements were identified by the two methods. In this way, all those with 70% or more points in common were considered as routine displacements, which was evidenced in the visual analysis of the displacements. Figure 2 shows the trajectories of an individual participating in the experiment in a given period that has at least 70% or more of their coincident points. It is visually evident that there is a trajectory that even partially contained in a frequent trajectory has more than 30% of its external points to it.

IV. RESULTS

During the survey, we identified 15 individuals who maintained a certain constancy in recording their movements in the period corresponding to the month prior to the XXIX Olympic Games, Beijing, 2008 and July 2009, the month following the record in the recorded pollution index 19 / 06/2009. Visual identification was performed using the Qgis software, where individual movements were plotted.

A. XXIX Olympic Games, Beijing, 2008

During the Olympic Games the point that most caught our attention were the trajectories of 6 individuals that were only realized on the days of the Games and did not have another record in their trajectories carried out in no other period. These trajectories were concentrated in places where some or several modalities were being performed and the Olympic Village (POI), indicating that the individuals changed their frequent trajectories to follow these events. Figure 3 shows the trajectories of each individual, identified with distinct colors, and the Olympic's POI.

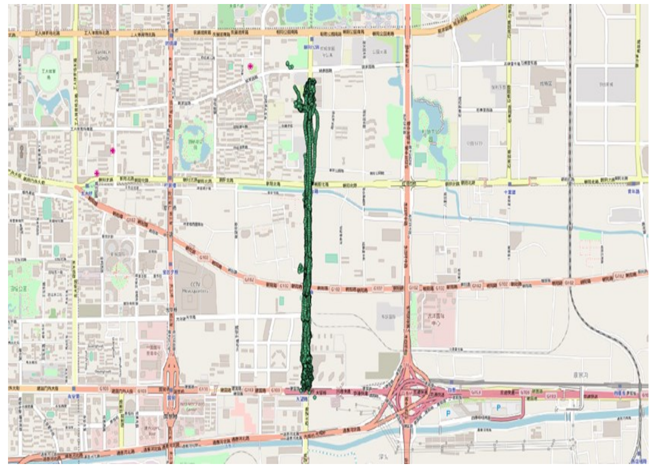

Figure 2 - Visual analysis of displacements

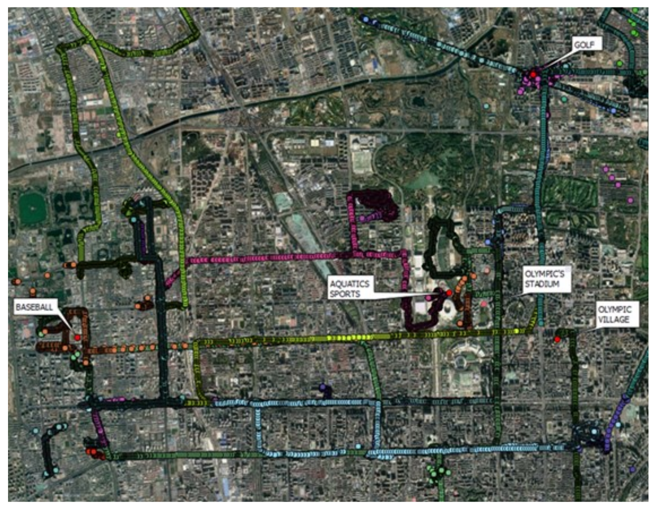

Figure 3 - Trajectories of several individuals

B. Peak Air Pollution

On June 19, 2009, the PM 2.5 index registered the peak of 712µg / m3 of air pollution, the highest of the year in Beijing. As can be seen from the table in Figure 4, this value indicates that it is possible to cause serious damage to health, even if exposure to the external environment is avoided.

AQI	Air Pollution Level	Health Implications	Cautionary Statement (for PM2.5)
0 - 50	Good	Air quality is considered satisfactory, and air pollution poses little or no risk	None
51 -100	Moderate	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.	Active children and adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion.
101-150	Unhealthy for Sensitive Groups	Members of sensitive groups may experience health effects. The general public is not likely to be affected.	Active children and adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion.
151-200	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects	Active children and adults, and people with respiratory disease, such as asthma, should avoid prolonged outdoor exertion; everyone else, especially children, should limit prolonged outdoor exertion
201-300	Very Unhealthy	Health warnings of emergency conditions. The entire population is more likely to be affected.	Active children and adults, and people with respiratory disease, such as asthma, should avoid all outdoor exertion; everyone else, especially children, should limit outdoor exertion.
300+	Hazardous	Health alert: everyone may experience more serious health effects	Everyone should avoid all outdoor exertion

Figure 4 - Air Quality Indexes

Having this date as a reference, the frequent trajectories and the changes that occurred in them on this day were identified. In the visualization, routine displacements were plotted as lines in green. The displacement on the day of the record pollution index was plotted as a red line.

Figure 5 shows that there is no significant change in the displacement on that day in relation to routine displacements. However, it is evident that the displacement was smaller in relation to the norm. Figure 6 this reduction of displacement is even more explicit.

Figure 5 - Urban mobility

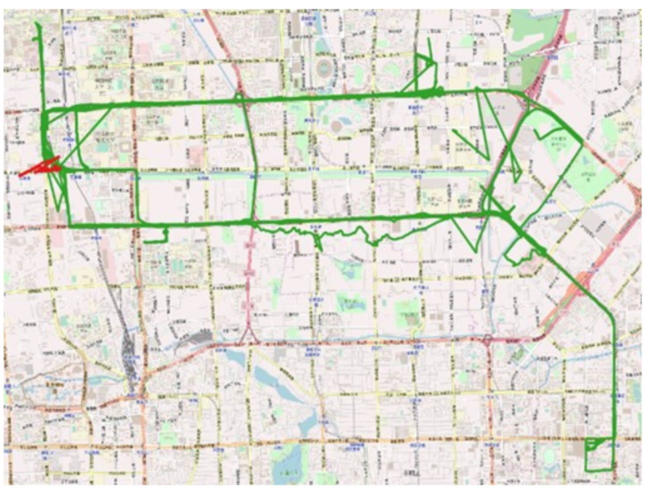

Figure 6 - Urban mobility

Figures 7, 8, 9 and 10 on the other hand show a noticeable change on said day. It should be noted that the most distant points identified in these displacements are in regions near mountains or areas with large amount of trees such as parks and clubs.

V. CONCLUSION

Adequate transport and circulation policies aimed at improving urban mobility should be related to systematic planning and to ecologically sustainable infrastructures in order to produce effective benefits. When planning for large-scale events this is especially important, since the collective behavior at special events changes the pattern of urban mobility causing convergence of movement to and from specific locations. Side effects, for example: increased pollution rates and locomotion difficulties caused by traffic congestion can be circumvented through the adoption of mobility strategies.

Unexpected events, climate or environmental changes (e.g. fires, snowstorms, pollution peaks, etc.) also alter urban mobility patterns and should be the target of urban planning policies. Mobile computing allows people to move and interact with information, services, and each other, resulting in a large amount of GPS trajectories that forms a history of people's location. Data on urban mobility, air quality and city dynamics have become widely available.

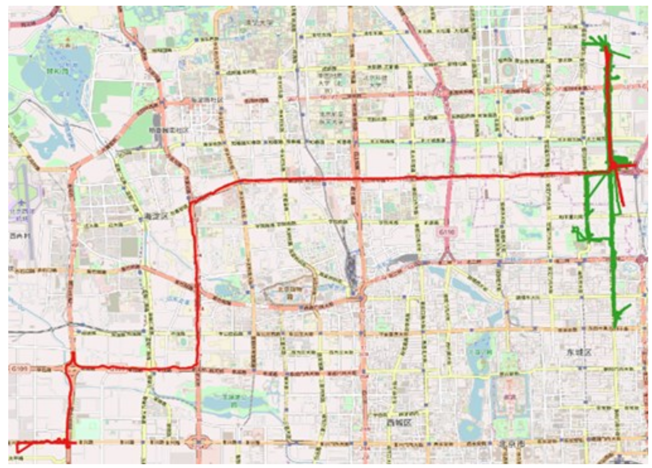

Figure 7 - Urban mobility

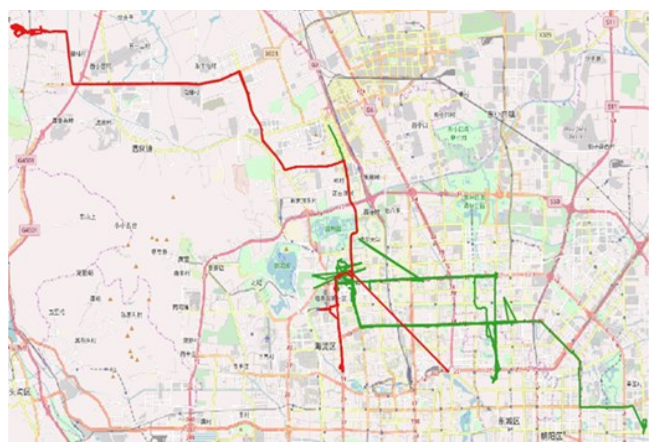

Figure 8 - Urban mobility

Moreover, monitoring these dynamics in real time opens up new possibilities for real-time response planning and redesign to handle any unforeseen effects or unexpected events. This goes beyond past data analysis for planning to enable on the spot decision making to support mobility.

The use of Artificial Intelligence techniques and Big Data processing on these data can not only be used in this planning, but also have great benefits in formulating strategies to avoid problems caused by events of great magnitude to react to these problems.

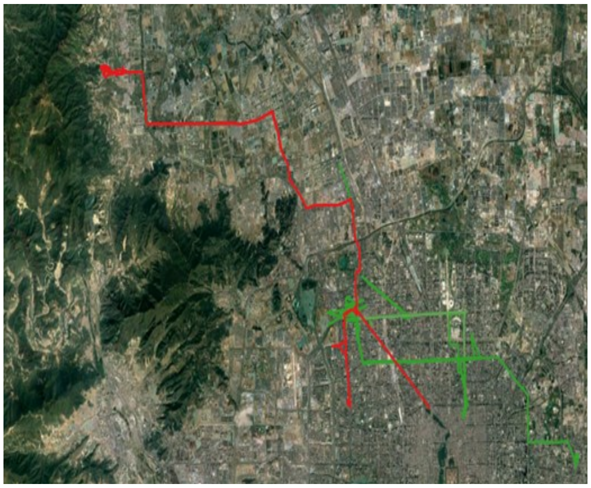

Figure 9 - Urban mobility

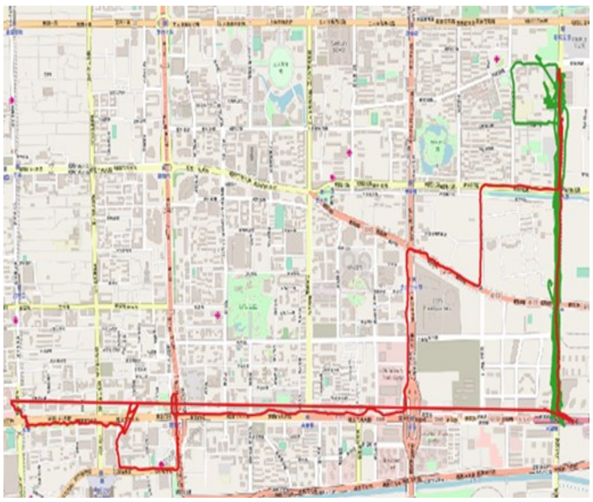

Figure 10 - Urban mobility

ACKNOWLEDGMENT

This research was supported by grants from CAPES, CNPq and FAPERJ.

REFERENCES

[1] Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of International conference on World Wild Web (WWW 2009), Madrid Spain. ACM Press: 791-800.

[2] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, Wei-Ying Ma. Understanding Mobility Based on GPS Data. In Proceedings of ACM conference on Ubiquitous Computing (UbiComp 2008), Seoul, Korea. ACM Press: 312-321.

[3] Yu Zheng, Xing Xie, Wei-Ying Ma, 2010. GeoLife: A Collaborative Social Networking Service among User, location and trajectory. Invited article, in the Data Engineering Bulletin of the IEEE. 33, 2, 2010, pp. 32-40.

[4] Zheng, Y., Liu, F. and Hsieh, H.P., 2013, August. U-air: When urban air quality inference meets big data. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1436-1444). ACM.

[5] Zhao, X., Xu, T., Fu, Y., Chen, E. and Guo, H., 2017, November. Incorporating Spatio-Temporal Smoothness for Air Quality Inference. In 2017 IEEE International Conference on Data Mining (ICDM) (pp. 1177-1182). IEEE.

[6] Zheng, Y., 2015. Trajectory data mining: An overview. ACM Trans. Intell. Syst. Technol. 6, 3, Article 29 (May 2015), 41 pages. DOI: http://dx.doi.org/10.1145/2743025

[7] Jiang, S., Ferreira, J. and Gonzalez, M.C., 2017. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. IEEE Transactions on Big Data, 3(2), pp.208-219.

[8] Kang, C., Sobolevsky, S., Liu, Y. and Ratti, C., 2013, August. Exploring human movements in Singapore: a comparative analysis based on mobile phone and taxicab usages. In Proceedings of the 2nd ACM SIGKDD international workshop on urban computing (p. 1). ACM.

[9] Xia, F., Wang, J., Kong, X., Wang, Z., Li, J. and Liu, C., 2018. Exploring human mobility patterns in urban scenarios: A trajectory data perspective. IEEE Communications Magazine, 56(3), pp.142-149.

[10] Kim, M., Kotz, D. and Kim, S., 2006, April. Extracting a mobility model from real user traces. In INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings (pp. 1-13). IEEE.

[11] Liang, Y., Zhou, X., Guo, B. and Yu, Z., 2012, December. Understanding the regularity and variability of human mobility from geo-trajectory. In Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01 (pp. 409-414). IEEE Computer Society.

[12] Juan Boubeta-Puig, Mario Bravetti, Luis Llana & Mercedes G. Merayo (2017) Analysis of temporal complex events in sensor networks, Journal of Information and Telecommunication, 1:3, 273-289

[13] Apparicio, P., Riva, M. and Séguin, A.M., 2015. A comparison of two methods for classifying trajectories: A case study on neighbourhood poverty at the intra-metropolitan level in Montreal. Cybergeo. 2015. 10.4000/cybergeo.27

[14] Huang, J., & Xiao, M. (2018). State of the art on road traffic sensing and learning based on mobile user network log data. Neurocomputing, 278, 110-118.

[15] Rifai, H., Kelly, P., Shoji, Y., Berry, D., & Zallio, M. (2018, January). Human Activity Detection Patterns: A Pilot Study for Unobtrusive Discovery of Daily Working Routine. In International Conference on Intelligent Human Systems Integration (pp. 143-148). Springer, Cham.