Visual Analytics of Traffic-Related Open Data and VGI

Jan Ježek*, Karel Jedlička**, Jan Martološ***

* University of West Bohemia in Pilsen/Department of Computer Science and Engineering, Pilsen, Czech Republic
** University of West Bohemia in Pilsen/Department of Mathematics, Pilsen, Czech Republic
*** EDIP s.r.o, Pilsen, Czech Republic
jezekjan@kiv.zcu.cz, smrcek@kma.zcu.cz, martolos@edip.cz

***Abstract*—Automobile traffic problems such as car accidents and traffic congestions are touching the daily life of many people. The development of demography and the urban planning play a key role that influences the traffic situation. An easily understandable and widely accessible visualization of current and future traffic volumes can influence nowadays decisions and bring important new insights.**

Currently available Open data together with VGI (volunteered geographic information) might be considered for the traffic volume prediction. In this paper, we focus on the utility of available Open data and VGI (e.g. OpenStreetMap[1]) together with advanced, web-based visualization technique (based on WebGL) with the aim to offer an easy exploration and insights discovery in complex data that relates to traffic. A particular focus is given to traffic volume and movement history analysis.

I. Introduction

Good prediction of a future traffic volume and density is important information for many subjects. The traffic is influenced by various aspects such as road capacity, traffic day-time variations, location of most visited places and events, actual and planned road constructions, weather conditions etc. These aspects make the gathering of relevant data a demanding and expensive process. Furthermore, an interactive visualization of an impact of a planned road construction, or future urban planning decision should be highly beneficial.

Nowadays there are many relevant data sources available on European level in the form of Open data, that might be used for traffic analytics and prediction (see e.g. Transmodel[2], the GTFS[3] and its applications[4] or DATEXII[5]) . Furthermore, there are new steadily developing web-based visualization approaches and tools that can provide advanced interactive techniques dedicated to exploratory visualization of complex problems that are related to traffic. Even though such tools and data are available for minimal costs in the public domain, their utility for such a purpose have not yet become a common practice.

In this paper we present a process of mining relevant parts of available Open data and VGI data sources that are suitable for prediction and history analysis of traffic. In particular we focus on prediction of traffic volume (the amount of vehicles which go through a network segment in a time period) by using OpenStreetMap data together with locally and globally available demographic data such as Eurostat. We also investigate into the analysis of historical traffic data by exploring GPS tracklogs, that are provided alongside with the OpenStreetMap. The main contribution of this paper is a workflow description dedicated to harmonization and processing of relevant Open Data and VGI in order to derive added value information about various aspects of traffic. Furthermore we demonstrate a novel web-based interactive visualization technique for traffic data by applying the concept of multiple coordinate views. Such a contribution is based on a intermediate results of OpenTransportNet project.

Next chapter summarizes the related work from the field of traffic analysis as well as visualization of multivariate spatio-temporal data. Section 3 describes the overall concept of prediction of traffic volume. Section 4 describes in details the data harmonization and data processing. Section 5 outlines the visualization techniques and depicts the final user interface. Finally the conclusion is given and a future work is mentioned.

II. Related Work

Live traffic information as well as a traffic prediction are important data for many subjects. A number of different methods for (automobile) traffic forecasting have been gradually developed. These methods can be divided into:

- trend methods (analogous) which assume that the prospective volume of traffic can be derived by extrapolation of current developments - basic principles described in [1].
- synthetic methods. which are based on examining patterns in the behavior of participants in the transport process and these principles also applied for a prospective period - see [2, 3].

Prognostic methods can be divided in relation to the treated area into two basic groups. These groups are:

- method of uniform growth factor, which assumes homogeneous development of the transport characteristics for the entire territory,
- mathematical model of transport network which calculate with local differences in land use - see [4].

Visualization of the history of traffic data has been explored by many different approaches. According to [5] they can be classified as: a direct visualization of raw data, a visualization of aggregated data and an extraction of derived information and its visualization. The [6] describes all parts of the problem of data processing and visual analytics of moving object data and mention various technical challenges that are faced during these tasks. One of the studies [7] describes an algorithm for an extraction of a specific points from moving object

[1] http://www.opentransportnet.eu/
[2] http://www.transmodel.org/index.html
[3] https://developers.google.com/transit/gtfs/
[4] https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds
[5] http://www.datex2.eu/

trajectories and offers a visualization technique for building a flow map. Another important aspect of discovering the semantic of the data is described in [8]. Such algorithms are usually designed to be performed upon request inside the dedicated database and provides typically aggregated values suitable for a visualization without the capabilities of live interaction.

One of the steadily developing topic of exploratory data visualization is the technique based on multiple coordinated views (MCV) given by [9]. Such a technique uses various visualization metaphors for different data types, where each visualization enables interaction (such as a filtering) that is further coordinated with an other view (e.g., selecting a time interval in a bar chart triggers immediate highlighting of relevant items in a map view). For the purpose of MCV visualization several frameworks a techniques exist [10, 11, 12]. In the paper we focus on the utility of web-based visualization technique for the purpose visualization of traffic volume and traffic history.

III. TRAFFIC VOLUMES

In the scope of this paper we target on two particular tasks: analysis and visualization of predicted traffic volumes and analysis and visualization of track logs provided by moving objects (e.g. cars). These task are based on scenarios collected during the OpenTransportNet as described by [13].

As it has been mentioned earlier the traffic volume is defined as a parameter of a road network which describes the amount of vehicles which go through a network segment in a time period. We can distinguish three types of traffic volumes:

- daily traffic volume (different for each day of the week),
- annual average of daily traffic volume (AADT),
- peak hour traffic volume – in the busiest hour of the day.

In general, there are three basic types of data necessary for traffic volume calculation using a mathematical traffic mode:

- Traffic generators - demographic data about places that are usually represented as points. These points can be cities, urban districts or building blocks – it depends on the granularity of the data and the desired level of detail. These data are used for estimation of traffic flows in the network. Distinguishing between different types of places such as living, industrial, service or shopping place is useful for estimation of traffic flows direction changes in time.
- Road network - well defined and topologically correct road network is the fundamental constraining graph structure, which describes allowed movements between different places.
- Calibration measurements - physical measurements of traffic volumes (traffic censes) at particular spots of the traffic network are used for calibration of calculated volumes.

To customize and process the relevant traffic data we have collected and downloaded all relevant data sources into one common database. For such a purpose PostgreSQL with PostGIS extension database server has been set-up. The process of mining and extracting relevant data from available Open Data sources is described in upcoming sections.

A. Traffic generators

The volume of the generated traffic can be determined as a function of the parameter characterizing the degree of attractiveness of the area for transport. This is due to the manner and degree of land use - the population in residential areas, the number of manufacturing employees, selling space in shopping areas.

Some of these data are freely available (population - Czech Statistical Office or EUROSTAT at the European level), some data are available for individual municipalities or regions in the land use planning documentation.

B. Route network construction

For the model of route network various data sources might be available, however different semantics are usually used. Therefore, we build a common schema based on the semantics used in the INSPIRE directive. Afterwards we derive the mapping from the source data to the INSPIRE data model. In the section we describe the harmonization of OpenStreetMap data into that model. The harmonization principles are similar for other regional data models as well.

The OSM data model (source model) was analysed and various data exports (namely geofabrik.de, OSM2PO and raw XML export) can be used. The OSM2PO tool, which is a PostGIS extension, was used for extracting source OSM data.

C. Common schema and harmonisation

The INSPIRE Transport Networks data model was chosen as the harmonised data schema, because it addresses the linear topology and is compliant with the EU legislation. The INSPIRE Transportation Networks schema was analyzed and then data structures necessary for routing (RoadLink and RoadNode) were selected. Then the mapping function from OSM data to the INSPIRE Transportation Networks schema was built. For such a purpose a mapping table has been designed, where we map the OSM-based code list to Inspire-based code list. The Inspire FunctionalRoadClassValue and FormOfWayValue code list mapped to OSM are depicted in Table 1 and Table 2.

TABLE I. CONVERT TABLE FOR FUNCTIONALROADCLASSVALUE

Inspire	OSM
FunctionalRoadClass Value	OSM.roads.type
mainRoad	motorway, motorway_link, trunk, trunk_link
firstClass	primary, primary_link
secondClass	secondary, secondary_link
thirdClass	tertiary, tertiary_link
fourthClass	Residential, living_street, unclassified
fifthClass	all other values

TABLE II. COVERSION TABLE FOR FORMOFWAY .

Inspire	OSM
FormOfWayValue	OSM.roads.type
bicycleRoad	cycleway
dualCarriageway	motorway_link, trunk, trunk_link, primary_link, secondary_link, tertialy_link

enclosedTrafficArea	raceway
entranceOrExitCarPark	not a corresponding value
entranceOrExitService	not a corresponding value
freeway	not a corresponding value
pedestrianZone	not a corresponding value
motorway	motorway
roundabout	not a corresponding value
serviceRoad	not a corresponding value
slipRoad	not a corresponding value
singleCarriageway	all other values
tractor	not a corresponding value
trafficSquare	not a corresponding value
walkway	pedestrian, footway, steps, path

The data harmonisation process was done in two steps: Already mentioned import of routing data from OSM to PostGIS database was done by the OSM2PO. Afterwards the data were converted into the INSPIRE-based database schema using PL/pgSQL functions. The result is the physical schema of spatial database stored in PostGIS and depicted in Figure 1.

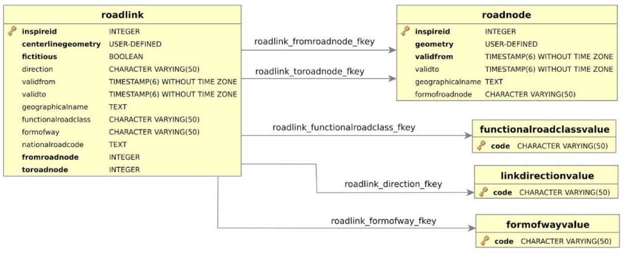

Figure 1. Physical schema of database

IV. History of Moving Objects

Another related task is to analyse and validate the predicted values by using real-world data. Ideal source for validation of traffic volumes are the data containing exact information about cars movement (GPS tracklogs - sequence of GPS coordinate measurements), however gathering such data in relevant temporal and spatial density is not an easy task. Even though such a data are collected by various corporations focused e.g. on the mobile phones industry (e.g. Google Inc.), to our best knowledge there are no Open Data with the appropriate density available for such a purpose.

A. GPS tracklogs as VGI

For the demonstration purposes, we use the GPS tracklog provided alongside with OpenStreetMap as a part of Planet.gpx initiative. Such a database offers millions of gps tracklogs collected by volunteers from all over the word. These data consist of tracklogs provided by pedestrians, bicycles, cars as well as airplanes, ships and other types of transportation. For our demonstration purpose we extract the subset of the data that is supposed to be provided by cars. In particular we extracted the tracklogs with average speed higher than 30km/h and lower than 150 km/h and with the track length greater than 5 km. Finally we choose a time period around May 2012 and Germany as region of interest, as we found it as one of the most appropriately covered by these tracks. We extracted records consisting of these items from the database:

- Moving object position expressed as the longitude and the latitude in a geographic coordinate system.
- Timestamps of such a position.
- Id of the moving object.
- Speed of the moving object.
- Type of the road additional derived from OpenStreetMap data.

V. Data Visualization

The data are used in two use cases (traffic volumes calculation and mapping of movement history), therefore two visualization techniques which communicate the results are presented in this chapter.

A. Traffic volume visualization

Traffic volume is a spatiotemporal feature with high dynamics in time. For demonstration of its dynamics, there was created an application during the Open Data Hackaton in Jelgava (September 2014) which serves as a proof of concept. See Figure 2 for a screenshot or even url http://gis.zcu.cz/projekty/OTN/TrafficVolumesExample.html for the live example. The width of the RoadLink shows the amount of vehicles crossing the segment per hour. The color shows, how close is the TrafficVolume to the maximum RoadLink capacity (green = 0 % - 50 %, yellow = 50 - 70 %, red = more than 70 %. Note also the time slider, which allows the user to see the data in various times.

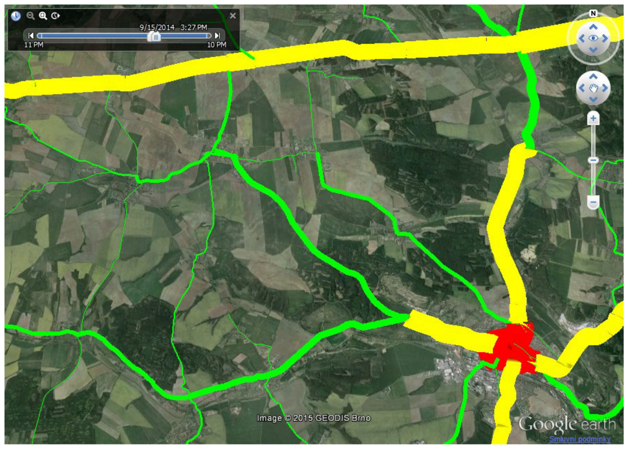

Figure 2. Traffic volume visualization

B. Visualization of GPS tracklogs

For a specified dataset we implemented a web-based multiple coordinated view visualization by using the WebGLayer and D3.js javascript libraries. In particular we configured these views:

- The number of GPS positions grouped by the rounded value of the speed displayed in the form of histogram.
- The number of GPS positions grouped by the hour of the day as a histogram
- The exact GPS position visualized as a symbol map. Symbol map visualize each position as a small square, where the color is used to express the speed. To overcome the problem of

overplotting a color transparency and blending is used.

The visualization is depicted in Figure 3. The Map is showing the positions and highlights those that are complaint with actual filters in the other views. The histograms on the right part shows the speed distribution and time of the day. The orange bars corresponds to the filter intersection, dark blue bars to the rest of the data visible in the map view, and the light blue bars to the data outside of the actual map window.

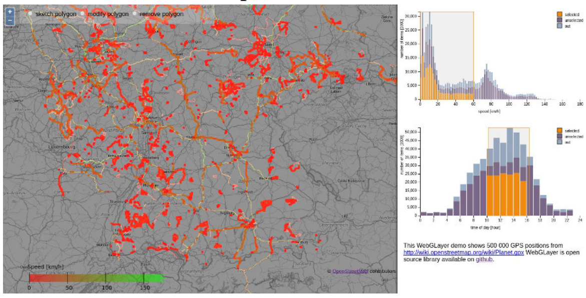

Figure 3. GPS Tracklogs visualization

The visualization is avalible as a demo of WebGLayer library at http://jezekjan.github.io/webglayer/.

VI. Conclusion

In this paper we demonstrate a combination of available open data and VGI for the purpose of car traffic analysis. For the future work we will apply mentioned methods in pilot cities of OTN project. Furthermore we target on users involvement in OTN pilot cities to trigger the interest to collect more accurate and actual VGI data. We wold like to further investigate into the more appropriate and accurate local data sources such as live feeds and on-line information about the traffic available in the form of the Open Data. We would like also to further enrich the visualization components by additional features such as visual comparison of regions of interest.

Acknowledgment

The authors of this paper are supported by the European Union's Competitiveness and Innovation Framework Programme under grant agreement no. 620533, the OpenTransportNet project.

This action is realized by the project EXLIZ – CZ.1.07/2.3.00/30.0013, which is co-financed by the European Social Fund and the state budget of the Czech Republic.

References

[1] Benjamin, Julian. "A time-series forecast of average daily traffic volume." Transportation Research Part A: General 20.1 (1986): 51-60.

[2] Főrschner, G., "Spezifisches Verkehrsaufkommen - eine wichtige Kenngrosse fűr die Strassenverkehrsplanung," Die Strasse 10, 1/70. Berlin 1970

[3] Lerman, Steven R.,"Location, housing, automobile ownership, and mode to work: a joint choice model," Transportation Research Record 610 (1976).

[4] Florian, M., editor, "Transportation Planning Models" North Holland, Amsterdam, The Netherlands (1984).

[5] Andrienko, G., Andrienko, N., "A visual analytics approach to exploration of large amounts of movement data", Springer Berlin Heidelberg, 2008.

[6] Andrienko, G., Andrienko, N., Bak, P., Keim, D., & Wrobel, S. (2013). Visual analytics of movement. Springer Science & Business Media.

[7] Andrienko, N., & Andrienko, G. (2011), "Spatial generalization and aggregation of massive movement data", Visualization and Computer Graphics, IEEE Transactions on, 17(2), 205-219.

[8] Bogorny, V., Renso, C., Aquino, A. R., Lucca Siqueira, F., & Alvares, L. O., "CONSTAnT–a conceptual data model for semantic trajectories of moving objects." Transactions in GIS, 18(1), 66-88. Chicago, 2014.

[9] Roberts, J. C. (2007, July). State of the art: Coordinated & multiple views in exploratory visualization. In Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on (pp. 61-71). IEEE.

[10] Ježek, J., Bernard, J., Kolingerová, I.: "WebGLayer for Advanced Spatial Data Exploratory Visualization on the Web", submitted to Goinformatica, Springer, 2015.

[11] Lins, L., Klosowski, J. T., & Scheidegger, C. (2013). Nanocubes for real-time exploration of spatiotemporal datasets. Visualization and Computer Graphics, IEEE Transactions on, 19(12), 2456-2465.

[12] Liu, Z., Jiang, B., & Heer, J. (2013, June). imMens: Real-time Visual Querying of Big Data. In Computer Graphics Forum (Vol. 32, No. 3pt4, pp. 421-430). Blackwell Publishing Ltd.

[13] Kozhukh, D., Jedlička, K., Mildorf, M.,Charvát, K.,Charvát K., Jr., Martolos, J., Šťastný, J., Traffic Volumes Described on Examples in the Open Transport Net Project Pilot Regions . Submitted to Proceedings of the 18th International Conference on Information Systems for Agriculture and Forestry (ISAF) 2014, Jelgava, Latvia, ISBN 978-80-905151-2-3 (under review).