Detection of potential updates of authoritative spatial databases by fusion of Volunteered Geographical Information from different sources

Stefan S. Ivanović[*1], Ana-Maria Olteanu Raimond[1], Sébastien Mustière[1], Thomas Devogele[2]

[1] Université Paris-Est, IGN, COGIT Lab, France
[2] Université François Rabelais de Tours, France

A continuous update of authoritative spatial databases is highly demanding task in both aspects, technical and financial. In the same time, alternative modalities to collect content, in particular spatial content, have achieved a certain maturity and must be considered as they may leverage the cost of updating authoritative spatial databases (Al-Bakri, 2010). This alternative data known as Volunteered Geographical Information – VGI (Goodchild, 2007) is easy available and is being collected in almost every moment somewhere in the world.

GPS tracks, in particular, seem to be a relevant source of update information to improve the freshness of a road network. Walkway, tractor and bicycle roads are identified as very challenging types for continuous update due to their intermittent nature (e.g. they appear and disappear very often) as well as various landscape (e.g. forest, high mountains, seashore, etc.). Even though, these types of roads are not of the highest priority for a national mapping agency (not a lot of resources are devoted to their update), they are still very important for production of touristic maps and for other different applications such as defense, sport activities, etc. Furthermore, a connectivity of the entire network depends on them also. The main objective of this research is proposing a method for identifying potential updates of mentioned road type of authoritative spatial databases using VGI data, more precisely GPS tracks.

Hence, we have focused on GPS traces obtained in sport activities, since they are mainly collected along those roads. Moreover, they are widely available on the websites of French sport association activities like RandoGPS, OpenRunner, UtagawaVTT etc.

 In: A. Comber, B. Bucher, S. Ivanovic (eds.): Proceedings of the 3rd AGILE Phd School, Champs sur Marne, France, 15-17-September-2015, published at http://ceur-ws.org

In order to detect potential updates - the differences between authoritative and VGI road networks, a data matching process is supposed to be applied. Matching links will be created between same roads represented in two different datasets (road networks).

As a result, three main situations were distinguished:

1. There is VGI trace but no corresponding IGN – a trace without matching link

2. There is IGN trace but no corresponding VGI – a trace without matching link

3. There are both, IGN and VGI trace – a trace without matching link

The first situation is related to the creation of a new road in the real world, which has not been represented yet in a corresponding authoritative spatial database. That is considered an alert for update and requires adding the road in the database.

The second situation represents a difference between authoritative and VGI data set in the way that a road which exists in a database, does not exist anymore in the reality. Subsequently, that road needs to be deleted from the database.

The third possible case is not a real alert for update, since there are no differences in compared data sets. However it is useful as a confirmation of a presence of the roads contained in the database in the reality.

However in the situations when there is more than one link either in VGI or authoritative dataset, we intend to estimate the average geometry of the trace and continue our work based on three situations described above.

Then, the question of VGI tracks quality arises. Furthermore, VGI traces are collected without any specified procedures, less or inexistent metadata, usually by low class GPS devices. Hence, heterogeneity of data is very high as well as spatial inaccuracy. In current stage of our work we focus on examination of data quality, especially on its spatial and temporal aspects.

First, we present an overview of VGI data sources (websites) and the heterogeneities that characterize them. In terms of data, I can rely on spatiotemporal data (i.e. coordinates and sometimes elevation and timestamps) as well as on a variety of descriptive information in text format such as: type of activity, difficulty, trace description etc.

To make the most of contextual information, we perform a comprehensive analysis of context elements which affects GPS data quality. Sources of errors related to technical aspect of GPS data collection are partially important for this research. Since we use data obtained by low class GPS receivers, which positional accuracy is at meter level, we are not concerned about the sources that affect the accuracy at sub-meter level. Therefore, our attention is directed to identifying and classifying sources of errors according to which extent they affect positional accuracy of GPS tracks.

Finally, we are interested in evaluation of data quality by analyzing VGI data itself – Intrinsic approach (Batini & Scannapieco, 2006). Thus, I tend to obtain the more statistical indicators of data quality that I can, such as indicators of: spatial dispersion, precision, reliability, correlation between data etc. As a result, a process of automatic collection of GPS traces from web-sites and storing them into PostgreSQL database was created, as well as a variety of tools designated to the indicators calculation. Evaluation of data quality is conducted by using an open source platform GeOxygene[1], developed by COGIT laboratory. Future work will aim to establishing a unique procedure for GPS tracks data quality evaluation.

References

- Goodchild, M.F. (2007). Citizens as sensors: The world of volunteered geography. GeoJournal, 69 (4), 211–221. doi:10.1007/s10708-007-9111-y.
- Al-Bakri, M.; Fairbairn, D (2010). Assessing the Accuracy of "Crowdsourced" Data and its Integration with Official Spatial Data Sets. In Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Leicester, UK, 20–23 July 2010; pp. 317–320.
- Batini, C., & Scannapieco, M. (2006). Data quality, Concepts, Methodologies and Techniques. In C. Batini & K. M. Scannapieco (Eds.), Approaches to the definition of Data Quality Dimensions: Empirical approach (pp. 39). Berlin Heidelberg, Germany: Springer-Verlag.

[1] Geoxygene : http://oxygene-project.sourceforge.net/