A FOSS4G-based procedure to compare OpenStreetMap and authoritative road network datasets

Maria Antonia Brovelli[1], Marco Minghini[1], Monia Molinari[1], Peter Mooney[2]

[1] Politecnico di Milano, Como Campus, DICA, Via Valleggio 11, 22100 Como, Italy

[2] Maynooth University, Department of Computer Science, Maynooth, Co. Kildare, Ireland

Abstract

OpenStreetMap (OSM), the most popular VGI crowdsourcing project, is an excellent example of an open-license spatial database. But what is the quality of OSM road network datasets compared to authoritative counterparts? Several comparisons of this kind are detailed in literature but these cannot be easily adapted to other scenarios. Developing a generic automated procedure is very challenging. This paper proposes a FOSS4G-based procedure for automated quality comparison of OSM and any authoritative road network datasets. We detail work-in-progress which has great potential. Our procedure is currently implemented into a GRASS command with future plans to extend this to a QGIS plugin and a FOSS4G-based WPS.

Keywords

Road network, FOSS4G, GRASS, open geo-data, OpenStreetMap, quality.

1 Introduction

Comparing open datasets is an interesting computational and societal challenge. Open geographic datasets from Volunteered Geographic Information (VGI) projects like OpenStreetMap (OSM) are excellent examples of crowdsourced geographic data of real world features such as buildings and roads. Recently some National Mapping Agencies (NMAs) have been making their geographic datasets available as open data. This greatly increases their distribution and opportunities for access for citizens. Comparing OSM data and open data from NMAs has been studied in the literature. There has been some work published reporting comparison of OSM and other official datasets displaying different levels of automation. Mooney, Corcoran, and Winstanley (2009) developed an automated quality assessment measure to compare OSM with authoritative datasets containing natural water features. Fairbairn and Al-Bakri (2013) compared OSM and authoritative large-scale databases in the UK and Iraq to address possible integration of these datasets. Ludwig, Voss, and Krause-Traudes (2011) developed a fully automated approach to matching street objects in Germany contained in OSM and in the commercial Navteq database. Haklay (2010) compared the OSM dataset and Ordnance Survey dataset in the UK, while Girres and Touya (2010) compared the quality of the

OSM dataset in France with the reference database from the French National Mapping Agency. Siebritz (2014) performed a quantitative and qualitative comparison between OSM and national mapping agency data in South Africa. In the studies outlined above (and some others in the literature) the authors have designed and developed software implementations of comparison methodologies which are specific and tightly coupled to the OSM and authoritative datasets involved. We believe that this makes the comparisons more complex and can prevent other authors from replicating them.

2 Our proposed comparison procedure

In this paper we propose a novel and generic procedure to perform comparisons of OSM and authoritative road network datasets in terms of spatial accuracy and completeness. The procedure is designed to a comparison framework which is general and applicable (in principle) to any pair of comparable open datasets while carefully considering their specific characteristics. When executing this procedure users can supervise the computation by manipulating parameter values to reflect the specific features of the authoritative dataset under investigation such as its scale and nominal accuracy. These characteristics make the procedure effectively adaptable to most comparisons of this kind.

2.1 Procedure implementation and workflow

The procedure is developed as a GRASS module written in Python. Its main steps are as follows: preliminary comparison of the datasets and computation of global statistics; geometric preprocessing of the OSM dataset to extract its networks representing the same road features as the authoritative dataset; and evaluation of the OSM dataset's spatial accuracy through a grid-based approach. These steps are separately described below by outlining their purpose, input/output data and GRASS modules used. REF (meaning “reference dataset”) is used to identify the authoritative road network dataset being compared with OSM.

2.1.1 Preliminary comparison of the datasets

The first step prepares the OSM and REF road datasets in addition to performing an initial comparison of their spatial coverage similarity. The key operations are as follows:

- import the OSM and REF datasets; if spatial clipping is required the user can import a vector layer to be used as the clipping mask (*v.in.ogr*, *v.overlay*);
- compute the total length of the OSM and REF datasets and their length difference, both in meters and percentages (*v.to.db*);
- apply a user-specified buffer around the REF and OSM datasets and compute the length and the length percentage of the OSM and REF datasets included in the buffer (*v.buffer*, *v.overlay*, *v.to.db*).

2.1.2 Geometric preprocessing of OSM dataset

This step prepares the OSM dataset so that only features which have a correspondence in the REF dataset are extracted. This correspondence is computed using angular coefficients as follows:

- apply a user-specified buffer around the REF dataset to extract only the OSM features included (*v.db.select*, *v.extract*, *v.buffer*, *v.overlay*);
- compute the angular coefficient of each feature in the REF dataset (*v.db.addcolumn*, *v.to.db*, *v.db.update*). This is compared to the angular coefficient of all OSM features falling inside the buffer around that REF feature. If the difference between the two angular coefficients exceeds a user-specified threshold (e.g. 30°), then that OSM feature is deleted as it does not correspond to the REF feature considered (see Figure 1). Corresponding OSM features are instead added into a new vector layer (*v.edit*, *v.patch*).

Before these operations the REF line features must be split into segments (*v.split*) allowing the angular coefficient be computed on each segment. This considerably increases the number of REF features for comparison and consequently the computational time required. For this reason the splitting operation is preceded by a generalization of the REF dataset (*v.generalize*) using the Douglas-Peucker algorithm (Douglas and Peucker, 1973). Users specify the threshold value for the line feature generalization (see Figure 1).

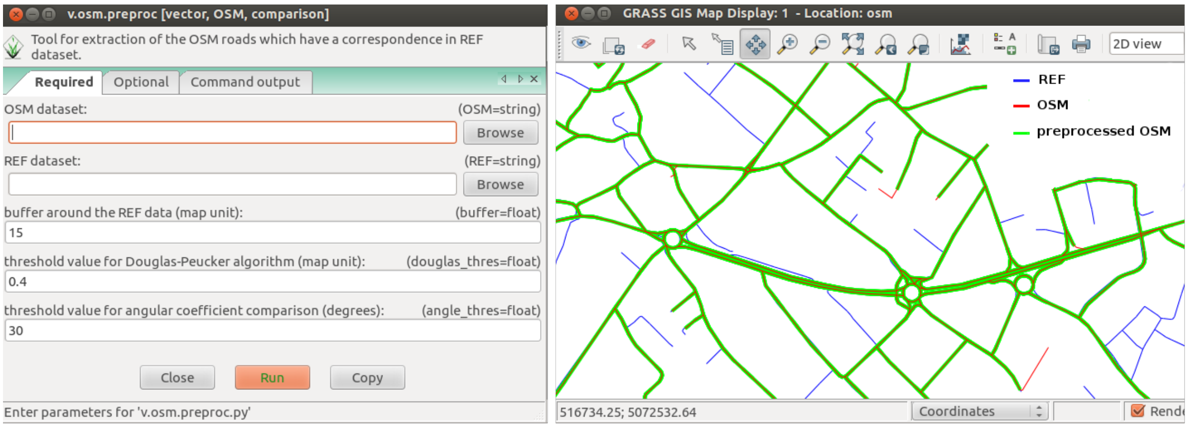

Figure 1: Geometric preprocessing of OSM dataset showing GRASS interface for user inputs (left) and REF, original OSM and preprocessed OSM roads (right).

2.1.3 Evaluation of OSM spatial accuracy using a grid-based approach

This step performs the comparison between the REF and the preprocessed OSM datasets and evaluates the spatial accuracy of the latter. This process is based on a grid approach which takes into account the possible heterogeneous nature (translating into an heterogeneous accuracy) of the OSM dataset. The single operations are the following:

- define a grid, asking the user to either build it in real time (by setting the bounding box and the grid step in the two directions (*v.mkgrid*)) or uploading a predefined polygon vector layer (*v.in.ogr*). The use of a grid is optional.
- apply one or more user-customized buffers around the REF dataset and compute the length and the length percentage of the OSM dataset included in the buffer (*v.buffer*, *v.overlay*, *v.to.db*).

The final operation is executed to obtain a number of different outputs according to user requirements. For example users can retrieve the grid cells where the deviation of OSM dataset from the REF dataset is lower than a fixed threshold or within a fixed interval of thresholds. Moreover they can evaluate which is the maximum deviation from the REF dataset for each grid cell.

3 Conclusions

The proposed procedure is under active development and research. We are working to achieve improved computational performances. In addition to the geometrical operations described here an analysis of the correctness of OSM road attributes (i.e. the values of the *highway* key) compared to the REF dataset is planned for future development. The procedure has been tested to compare the OSM road network dataset with those authoritative open datasets of the municipalities in the Lombardy Region in Northern Italy (scale 1:2000). Our immediate future work should, at first, confirm the suitability of the procedure – which, as shown, is heavily supervised by the user input values – on the input datasets. From a FOSS4G viewpoint the authors will develop this procedure to become a QGIS plugin before being provided as a Web Processing Service (WPS) available for the whole community.

Acknowledgements

The authors grateful acknowledge support from COST Action TD1202 “Mapping and the Citizen Sensor” and specifically the Short Term Scientific Mission (STSM) COST-STSM-TD1202-24885.

References

- ✔ Douglas, D. H., & Peucker, T. K. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *10*(2), 112-122.
- ✔ Fairbairn, D., & Al-Bakri, M. (2013). Using Geometric Properties to Evaluate Possible Integration of Authoritative and Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, 2(2), 349–370.
- ✔ Girres, J.-F., & Touya, G. (2010). Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS, 14*(4), 435–459.
- ✔ Haklay, M. (2010) How good is OpenStreetMap information? A comparative study of OpenStreetMap and Ordnance Survey datasets for London and the rest of England. *Environment and Planning B*, 37, 682-703.
- ✔ Ludwig, I., Voss, A., & Krause-Traudes, M. (2011). A Comparison of the Street Networks of Navteq and OSM in Germany. In S. Geertman, W. Reinhardt & F. Toppen (Eds.), *Advancing Geoinformation Science for a Changing World* (pp. 65-84). Springer Berlin Heidelberg.
- ✔ Mooney, P., Corcoran, P., & Winstanley, A. C. (2010). Towards Quality Metrics for OpenStreetMap. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (pp. 514–517). New York, NY, USA: ACM.
- ✔ Siebritz, L.-A. (2014). Assessing the accuracy of openstreetmap data in south africa for the purpose of integrating it with authoritative data. PhD Thesis, University of Cape Town, SA.