Towards Scalable Geospatial Remote Sensing for Efficient OSM Labeling

Rui Zhang[1], Marcus Freitag[1], Conrad Albrecht[1,*], Wei Zhang[1] and Siyuan Lu[1]

[1] Data Intensive Physical Analytics, TJ Watson Research Center, IBM Research, Yorktown Heights, NY, United States; rui.zhang@ibm.com, mfreitag@us.ibm.com, cmalbrec@us.ibm.com, weiz@us.ibm.com, lus@us.ibm.com

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2019 Conference in Heidelberg after peer-review.

The past decade has shown a dramatic increase in the amount of openly available geo-spatial datasets such as multi-spectral and RADAR satellite imagery from space agencies like ESA and NASA government-sponsored, high-resolution aerial survey raster data, e.g., USDA (U.S. Department of Agriculture) weather reanalysis model data based on sensor networks, e.g. published by the PRISM Climate group geo-tagged messages as well as images from social media platforms such as Twitter and Instagram, etc. accumulating geo-tagged information at data rates easily exceeding tens of terabytes a day.

Given that an open database project such as OpenStreetMap (OSM) relies on volunteers to spend their valuable time to generate vector datasets that annotate and update information on roads, buildings, land cover, points of interest, etc., it is natural to ask how sources of freely available spatio-temporal information might help to support and guide mappers in their work.

At the same time, major progress has been made in the open-source digital arena of big data processing and artificial intelligence (AI). For example, projects for distributed non-relational database systems such as HBase (https://hbase.apache.org/) or in-memory distributed compute frameworks such as Spark are available to run on commodity hardware to scale analytics. Deep learning libraries such as PyTorch (https://pytorch.org) in accordance with the explosive amount of neural network architectures published by academia enable, for example, state-of-the-art computer vision algorithms which can be leveraged for remote sensing tasks: detection of buildings, land classification, change detection, etc.

Our work discusses and demonstrates how to link tools from big data analytics and machine learning to geo-spatial datasets at scale in order to extract value from openly available spatio-temporal datasets to the potential benefit of OSM mappers. In particular, we show the design of a system that employs the key-value store HBase to index spatio-temporal satellite imagery to let Spark-SQL (https://spark.apache.org/sql/) user-defined functions act on it to remotely identify human signatures on Earth's surface by the aid of AI.

Zhang et al. (2019). Towards Scalable Geospatial Remote Sensing for Efficient OSM Labeling
In: Minghini, M., Grinberger, A.Y., Juhász, L., Yeboah, G., Mooney, P. (Eds.). Proceedings of the Academic Track at the State of the Map 2019, 27-28. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3387715

Finally, when it comes to pixel-wise land classification, we are using the 1-meter resolution USDA aerial survey data and information derived from the OpenStreetMap project. The goal is to establish a scalable pixel level translation model from aerial map to OSM, where colors and shapes define land classification, i.e., forestry, grassland, building, road, etc. The USDA aerial survey is refreshed every other year, so we expect to translate the latest aerial survey to OSM and compare with the current OSM state to identify changes on the actual land use. This information will guide the OSM community where the map needs to be updated.

We believe that the techniques and use cases presented will help to identify hot spots of where OSM needs human labor most – either in mapping or updating labels. Moreover, we hope to spark a scientific, strategic and technical conversation with the OSM community on needs regarding semi-automated support systems for global mapping. As a bonus, we will introduce the open-source tool [1] to interact with the spatio-temporal platform PAIRS that supports our research.

References

[1] IBM (2019). IBM PAIRS Geoscope. Retrieved from https://github.com/IBM/ibmpairs