Analytics of big geosocial media and crowdsourced data

Songnian Li, Monica Wachowicz & Hongchao Fan

To cite this article: Songnian Li, Monica Wachowicz & Hongchao Fan (2021) Analytics of big geosocial media and crowdsourced data, Big Earth Data, 5:1, 1-4, DOI: 10.1080/20964471.2021.1898780

To link to this article: https://doi.org/10.1080/20964471.2021.1898780

© 2021 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the CASEarth Strategic Priority Research Programme.

Published online: 06 Apr 2021.

Submit your article to this journal

Article views: 818

View related articles

View Crossmark data

Citing articles: 1 View citing articles
Analytics of big geosocial media and crowdsourced data

Numerous crowdsourcing and social media platforms such as CrowdSpring, Idea Bounty, DesignCrowd, Facebook, Twitter, Flickr, Weibo, WeChat, and Instagram are creating and sharing vast amounts of user-generated content that can reveal timely and useful information for detecting traffic patterns, mitigating security risks and other types of time-critical events, discovering social structures characteristics, predicting human movement, etc. Crowdsourcing, also known as volunteered geographic information (VGI), has added a new dimension to traditional geospatial data acquisition by providing fine-grained proxy data for human activity research in urban studies (Chen et al., 2016; Niu & Silva, 2020). However, analyzing big geosocial media and crowdsourced data brings significant methodological and theoretical challenges due to the uncertain user representability when referring to human behavior in general, the inherent noisy data that requires high-performance cost of preprocessing, and the heterogeneity in quality and quantity of sources. In particular, geosocial media data and their derived metrics can provide valuable insights and policy strategies, but they require a deep understanding of what the metrics actually measure (Zook, 2017). All of these underpin complex assessments, not mentioning the ethnic and privacy issues. Therefore, new sets of methods and tools are required to analyze the big data from crowdsourcing and social media platforms.

This special issue aims at a collection of the state-of-the-art research efforts that address these challenges, including techniques, practices, and applications in the integration of different types of big geosocial media and crowdsourced data for developing and evaluating methods and techniques for semantic classification of web services, mobility, 3D building modeling, data disaggregation, quality control, and visualization of transportation data. As illustrated in Figure 1, handling big crowdsourced data is an important fundamental part of the research studies reported in the included articles.

The special issue starts with a review article titled “A review of the use of geosocial media data in agent-based models for studying urban systems” by Richard Wen and Songnian Li. Through a review of recent agent-based modelling applications that incorporate geosocial media data in the context of urban systems, the authors explore how their integration can help urban studies and what challenges and opportunities remain. It concludes that the complex and dynamic nature of urban systems has made agent-based modelling an appropriate and effective approach for studying the social interactions in urban systems, while the integration of a sheer amount of geosocial media data with agent-based modelling can optimize ABMs, which can be applied to various real-world applications involving social infrastructures in urban systems.

With the emergence of the Internet of Moving Things (IoMT) as a new technological infrastructure for sensing and communication, Black and Wachowicz, in their article “Clustering spatio-temporal bi-partite graphs for finding crowdsourcing communities in IoMT networks”, describe a bi-partite graph based on representing the changes of mobility relationships. This graph representation is needed to identify a volunteer...
organization of neighbor devices that are moving close to one another in our physical world and seamlessly connected in their virtual world. The Louvain community detection method is proposed to find communities of intelligent devices to reveal a value conscious participation of citizens. The proposed bi-partite graph model is evaluated using a real-world scenario in transportation, confirming the main role of evolving communities in developing crowdsourcing IoMT networks.

In the next article on the “An Interactive platform for low-cost 3D building modeling from VGI data using convolutional neural network”, the authors Hongchao Fan, Gefei Kong, and Chaoquan Zhang used crowdsourcing data such as street-view and user-uploaded images for generating 3D building models. In their work, an interactive platform was developed based on WebGL technology that can effectively generate a 3D building model from images in 30 seconds, with the help of a user interaction module and a convolutional neural network (CNN). The user interaction module provides the boundary of building facades for 3D building modeling. Then the CNN detects facade elements despite multiple architectural styles and complex scenes. This work has the potential to become a powerful tool in a VGI project to collect 3D building models with semantic information.

Web mapping and the use of geospatial information online have rapidly evolved over the past few decades, yielding a proliferation of web map services (Veenendaal, Brovelli, & Li, 2017). However, discovering desired map resources from the massive web mapping services is still challenging due to the lack of standards and service registration. The article “Text GCN-SW-KNN: a novel collaborative training multi-label classification method for WMS application themes by considering geographic semantics” by Zhengyang Wei and others proposes a novel multi-label text classification method to facilitate the search of web mapping services based on their metadata. The classification method adopted two base models: an improved Text Graph Convolutional Network and a widely used MLKNN. It was tested using OGC WMSs acquired by a topic-focused web crawler.
The importance of data standardization and quality control of the viral case data from scattered sources has drawn researchers’ attention during the ongoing effort on combating the COVID-19 pandemic. The article “A spatiotemporal data collection of viral cases for COVID-19 rapid response” by Dexuan Sha and others develops a comprehensive data production workflow with a standardized spatiotemporal data format, accuracy data quality control, and consistent and timely operation. In their approach, data from countries and regions are crowdsourced from external sources with varying formats using customized data crawler scripts, processed and validated through a semi-automatic validation workflow, and stored as spatiotemporal data cubes. The final data products are released and shared over the Internet.

Data disaggregation has been fundamental for many geospatial applications, especially the full implementation of the sustainable development goal (SDG) indicator framework (IAEG-SDGs, 2019). In their article “A 100 m population grid in the CONUS by disaggregating census data with open-source Microsoft building footprints”, Huang and his collaborators describe an approach for generating a 100-meter population grid in the Contiguous United States (CONUS). The proposed approach consists of disaggregating the US census records using the Microsoft Building Footprints and the crowdsourced OSM land-use dataset from the OpenStreetMap (OSM). Dasymetric mapping is employed with weighting based on trimmed building footprints to disaggregate population data. In addition to proving the feasibility of using open and crowdsourcing data for population disaggregation, the study also generated disaggregated population data products that have been made publically available.

Multi-source and multi-form data, such as open, crowdsourced, or passively collected data, are crucial to many intelligent transportation system (ITS) applications. Due to their big data nature, one of the challenges is to be able to make such data accessible to end users for real-world decisions, where data visualization tools and techniques play a critical role in analyzing and interrogating the data by researchers, policy and decision-makers, and citizens (Li et al., 2016). The last article “The visual analytics of big, open public transport data – a framework and pipeline for monitoring system performance in Greater Sydney” by Oliver Lock, Tomasz Bednarz, and Christopher Pettit, explores visual analytics techniques using a passively-collected, real-time big data set obtained through open data feeds. Although not closely dealing with crowdsourcing data, the insights and lessons learned from studying this sheer amount of open data feeds should benefit in improving visual analytics of other types of big crowdsourced data.

We want to thank all contributing authors, including those whose papers were not selected for publication in this special issue. Special thanks go to the many anonymous reviewers. Without their contributions, this special issue would not have been published. Finally, we would like to extend our thanks to the Executive Editor-in-Chief Dr. Changlin Wang and Assistant Editor Ms. Linlin Guan for their assistance, guidance, and patience.

**Funding**

This work was supported by the Natural Sciences and Engineering Research Council of Canada [RGPIN-2017-05950].
References

Chen, J., Dowman, I., Li, S., Li, Z., Madden, M., Mills, J., . . . Heipke, C. (2016). Information from imagery: ISPRS scientific vision and research agenda. *ISPRS Journal of Photogrammetry and Remote Sensing, 115*, 3–21.

IAEG-SDGs. (2019, March 5-8). Data disaggregation and SDG indicators: policy priorities and current and future disaggregation plans. A Background document presented in Statistical Commission, Fiftieth session.

Li, S., Dragicevic, S., Anton, F., Sester, M., Winter, S., Coltekin, A., . . . Cheng, T. (2016). Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing, 115*, 119–133.

Niu, H., & Silva, E. A. (2020). Crowdsourced data mining for urban activity: Review of data sources, applications, and methods. *Journal of Urban Planning and Development, 146*(2), 04020007.

Veenendaal, B., Brovelli, M. A., & Li, S. (2017). Review of web mapping: Eras, trends and directions. *ISPRS International Journal of Geo-Information, 6*(10), 317.

Zook, M. (2017). Crowd-sourcing the smart city: Using big geosocial media metrics in urban governance. *Big Data & Society, 4*(1), 2053951717694384.

Songnian Li
*Ryerson University, Toronto, Canada*
✉ snli@ryerson.ca  🌐 http://orcid.org/0000-0002-8244-5681

Monica Wachowicz
*University of New Brunswick, Fredericton, Canada*
🌐 http://orcid.org/0000-0002-4659-0101

Hongchao Fan
*Norwegian University of Science and Technology, Trondheim, Norway*
🌐 http://orcid.org/0000-0002-0051-7451