Chapter 5
Improving Geospatial Big Data Analytics Approaches:
A Focus on High-Velocity Data Streams

Sana Rekik
Higher Institute of Management of Tunis, Tunisia

ABSTRACT

The advent of geospatial big data has led to a paradigm shift where most related applications became data driven, and therefore intensive in both data and computation. This revolution has covered most domains, namely the real-time systems such as web search engines, social networks, and tracking systems. These later are linked to the high-velocity feature, which characterizes the dynamism, the fast changing and moving data streams. Therefore, the response time and speed of such queries, along with the space complexity, are among data stream analysis system requirements, which still require improvements using sophisticated algorithms. In this vein, this chapter discusses new approaches that can reduce the complexity and costs in time and space while improving the efficiency and quality of responses of geospatial big data stream analysis to efficiently detect changes over time, conclude, and predict future events.

INTRODUCTION

Since its emergence, the big data have managed to stand as an important multidisciplinary phenomenon involving various disciplines and areas of research. These latter may include parallel and distributed systems, communication networks and data analysis fields. Big data as a research field has efficiently affected most of

existing domains including health, industry, environment and transportation (Lytras, 2017). Big data is defined as a massive amount (volume) of data that exceeds the limit of traditional storage powers. However, professionally, it is defined through a set of relevant characteristics named the Big data Vs, where the volume is just one of those features. These Vs define a Big data model that ranges from 3Vs to 7Vs (Uddin, 2014). Recently, the range has extended to reach 10Vs models (Khan, 2018). These Vs are respectively, the Volume of data, the Velocity that defines the high speed of data production and generation, the Variety that defines the heterogeneity of data, Value that refers the usefulness of the data, and the Veracity which refers to the quality of data (Storey, 2017). Equally, the other Vs are, Viability, Validity, Variability, Volatility, Viscosity, which are more explained in (Khan, 2018).

Due to its variety Big data includes different types of data, the Geospatial (geo) is an important one of them that often refers to geographical location. Apart the use of traditional capture tools, this geospatial data can be generated using geo-located sensors, Internet of Things objects, tracking systems and volunteered geographic information (VGI) sources such as social networks (Li, 2016). Geospatial data is applied in most of all applications as supply chain management (Irizarry, 2013), traffic and health monitoring services (Griffin, 2015). This has led to the appearance of a specific sub-domain; Geospatial big data. This later has demonstrated significant support for real-time systems through the application of several analytics approaches (Lee, 2015). Thus, researchers have opened new opportunities and focus to further leverage geospatial Big data in such context. On the other hand, given the complexity of geospatial data, related operations are becoming more and more complex and costly. Therefore, traditional geospatial tools, whether for storage, processing, analyzing, or visualization, have reached their limits facing these constraints (Robinson, 2017). Consequently, an emerging need has occurred to migrate to new geospatial big data tools and techniques.

Data-intensive applications have emerged last years and frequently appeared in various fields given the wide availability of data and also calculation techniques (Chen, 2014). This evolution required the study of the different challenges of these applications and generally of the big data phenomenon.

As already mentioned, high-velocity is among the geospatial Big data characteristics (Vs). This latter is a significant factor for stream-based applications, which requires real time responses and therefore a low latency for data stream processing and analysis. Therefore, processing real time data have more requirements than other, given that the time constraint and the speed of the response represent an important priority in such real-time applications. Therefore, such streams which data change frequently over time require sophisticated algorithms to identify these changes. In this context, researches have to be directed toward applying new thinking strategies and methods to cope with scaling problems. This principally aims to find a compromise without

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/improving-geospatial-big-data-analytics-approaches/279251

Related Content

Quality Evaluation of Volunteered Geographic Information: The Case of OpenStreetMap
Hongyu Zhangand Jacek Malczewski (2017). *Volunteered Geographic Information and the Future of Geospatial Data (pp. 19-46).*
www.irma-international.org/chapter/quality-evaluation-of-volunteered-geographic-information/178797

A Cross Reading of Landscape Through Digital Landscape Models: The Case of Southern Garda
Ilaria Forti (2019). *Geospatial Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 1144-1174).*
www.irma-international.org/chapter/a-cross-reading-of-landscape-through-digital-landscape-models/222941

Minimizing Construction Emissions Using Building Information Modeling and Decision-Making Techniques
Mohamed Marzoukand Eslam Mohammed Abdelkader (2017). *International Journal of 3-D Information Modeling (pp. 14-35).*
www.irma-international.org/article/minimizing-construction-emissions-using-building-information-modeling-and-decision-making-techniques/192121

Spatiotemporal Network Analysis and Visualization
Judith Gelernterand Kathleen M. Carley (2015). *International Journal of Applied Geospatial Research (pp. 77-97).*
www.irma-international.org/article/spatiotemporal-network-analysis-and-visualization/122363

Map Matching Algorithms for Intelligent Transport Systems
Mohammed A. Quddus (2009). *Handbook of Research on Geoinformatics (pp. 302-309).*
www.irma-international.org/chapter/map-matching-algorithms-intelligent-transport/20417