Spatial Data Mining: A Perspective of Big Data

Shuliang Wang, School of Software, Beijing Institute of Technology, Beijing, China
Hanning Yuan, School of Software, Beijing Institute of Technology, Beijing, China

ABSTRACT

Big data brings the opportunities and challenges into spatial data mining. In this paper, spatial big data mining is presented under the characteristics of geomatics and big data. First, spatial big data attracts much attention from the academic community, business industry, and administrative governments, for it is playing a primary role in addressing social, economic, and environmental issues of pressing importance. Second, humanity is submerged by spatial big data, such as much garbage, heavy pollution and its difficulties in utilization. Third, the value in spatial big data is dissected. As one of the fundamental resources, it may help people to recognize the world with population instead of sample, along with the potential effectiveness. Finally, knowledge discovery from spatial big data refers to the basic technologies to realize the value of big data, and relocate data assets. And the uncovered knowledge may be further transformed into data intelligences.

Keywords: Big Data, Data Intelligence, Geomatics, Spatial Big Data Mining

INTRODUCTION

Big data is a complex data set with the characteristics such as Volume, Variety, Velocity and Veracity (Meyer-Schoenberger, & Cukier, 2013). It exceeds the capacity of current computing techniques and tools (Wu et al., 2014; Rajaraman, & Ullman, 2011). In big data, spatial data specifically accounts for the vast majority. About 80% of data is associated with the spatial position (Grossner, Goodchild, & Clarke, 2008; Shekhar, Xiong, 2007). Any kind of object or structure may be related to a geo-spatial location, such as buildings, roads, places, bridges, or rivers. Spatial data is the basis of data and source of wisdom for people to understand the real-world through the information world by better addressing social, economic, and environmental issues of pressing importance (Hackeloer et al., 2014).

Big data is closely associated with various applications, in which spatial data mining is principal. Because of the complex types, intrinsic relationships, and implicit autocorrelation in spatial big data, it is much more difficult to extract the useful patterns from spatial datasets.
than from conventional numeric and categorical datasets (Vatsavai et al., 2012).

**From the Real World to the Information World**

Human civilization is a process from understanding the real world to creating the information world. It has gone through the following stages: information being sensed from the world, information assisting in memorizing the world, information being recorded and inherited, exchange and communication by information, and understanding the world once again by information (Li, Wang, & Li, 2013). Initially, human take advantage of stones and shells to count according to the principle of one-to-one. And they keep records by tying knots. Later, picture notes are invented to inherit more accurate memory by using simple graphics along with perceptual cue. When the picture becomes relatively fixed by shape symbols, associating with the words in the language, it produces texts. Texts abstract and generalize the world, promote cultural understanding, and prepare the necessary foundation for the development of science. In order to break through the restrictions that the written symbols depend on artificial transcription, carving or engraving, human use machines after industrial revolution to batch mechanization production, which promotes the efficiency of the cultural transmission. Computer pays attention to high-speed computing, and spins off the software from the hardware, contributing to the dissemination of information “electronically” and “automatically”. Internet focuses on network by interrelating computers, and it breaks local information limitation. Mobile communication unbounds human from the machine by making the machine follow user’s movements. The Internet of Things cares for the application of human-machine interoperation by automatically identifying objects, enabling the information sharing between the human and things. Cloud computing centralizes service by consolidating expertise and optimizing the allocation of resources. Big data aims at population data, and takes good advantage of the entire data, breaking the randomness of the samples (Burstein, Holsapple, 2008; Haining, 2003), and demonstrating on big data center and mobile terminal. The abovementioned information technologies ultimately serve for the understanding and transforming of the real world.

**More Attention to Big Data**

As the real world is explored through scientific research, humans also try to discover the mysteries in the information world through big data. The discovery attracts much attention from academia. In July 2008, O’Reilly Media published a monograph entitled “Beautiful Data”. In September, a special issue on “Big data” came out in “Nature”, and a book entitled “The Fourth Paradigm—Data Intensive Scientific Discovery” was published. In May 2009, the project “Global Pulse” of the United Nations released “Big data for Development: Challenges & Opportunities” for promoting the innovating ways to rapidly collect and analyze digital data (United Nations Global Pulse, 2012). In February 2011, “Science” published the special issue on “Dealing with Data” and joined its sub-journals as “Science: Signaling”, “Science: Translational Medicine” and “Science: Careers” to launch related topics, for discussing the importance of data for scientific research. In May, McKinsey Global Institute (2011) analyzed the potential application of big data in different industries from the economic and commercial dimensions, and spelled out the development policy on how the decision-makers dealt with big data in governmental and industrial organizations. In January 2012, the “Wall Street Journal” argued that information technology had entered a big-data era, and the coming tech-led boom would be big data, smart manufacturing and the wireless revolution (Mills, &Ottino, 2012). In March 2012, the United States government released “Big data Research and Development Initiative”, which rose the development and application of big data from business conduct to national deployment strategic in order to improve the ability to extract knowledge from
Data Mining in Designing an Agent-Based DSS
Christian Bohm, Maria R. Galli and Omar Chiotti (2003). Data Mining: Opportunities and Challenges (pp. 421-436).
www.igi-global.com/chapter/data-mining-designing-agent-based/7612?camid=4v1a

Electric Vehicles in the Smart Grid
www.igi-global.com/chapter/electric-vehicles-in-the-smart-grid/90078?camid=4v1a