BIG DATA AND ITS APPLICATIONS IN AGRO-GEOINFORMATICS

Liping Di

Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA

ABSTRACT

Agro-geoinformatics deals with collecting, managing, analyzing agricultural-related geospatial data, which are domain-specific big data. This paper discusses the general characteristics of big data, the specific features of agro-geoinformatics and agro-geodata, and examples of agro-geoinformatics projects dealing with big agro-geodata. Through the adoption and adaptation process, the general big data technologies are very useful in agro-geoinformatics, but cannot solve all technology needs in dealing with big agro-geodata. Development of agro-specific big data technology is a necessary supplement to adoption of general big data technology. The combination of adoption of general big data technology and development of agro-specific big data technology proves to be a good strategy for applying big data technology in agro-geoinformatics.

Index Terms— Big Data, Agriculture, Agro-geoinformatics, Agro-geodata

1. INTRODUCTION

With the rapid development of sensor, sensing, and data collecting technologies, the capabilities of human society to collect data have been expanded exponentially. Huge amount of data has been collected by government agencies, organizations, industries, and individuals. The development of data interoperability technology and wide adoption of open data policy have made the data more accessible with low or no cost. However, the data are very diverse in terms of sources, format, quality, and etc. The human society lucks of enough experience and knowledge in managing and exploring the rapidly increased volumes of data. As the result, most of those collected data are either discarded or put into achieves without being fully utilized, although the data may contain information and knowledge that are valuable to the socio-economic activities of the society.

2. BIG DATA

Realizing the socioeconomic values of information and knowledge contained in huge volumes of the data, in recent years significant research efforts have been spent on maximizing the utilization of the data by mining the information and knowledge from the data [1]. The term "**Big data"** is coined to refer to the huge amount of data those efforts deal with. According to Wikipedia, "Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate."[2]

Big data are commonly characterized with five Vs [3]:

- **Volume** refers to the huge amounts of data generated every day. For example, millions of cameras have been installed worldwide to monitor earth environment, traffic conditions, public safety, and etc. year around. The volume of data generated by those cameras is un-imaginable.
- **Velocity** refers to the speed at which data is generated and moved around. Every second the world generates petabytes of data, which need to be managed and analyzed, and near-real-time decision might be made based on the analysis results.
- **Variety** refers to the different types of data the world generates and uses. For example, in the geospatial field we now need to deal with data from in-situ, airborne, and satellite platforms, and from citizen scientists' mobile devices. The data type can range from hyper-spectral images, videos, model outputs, to social media conversations.
- **Veracity** refers to the trustworthiness of the data. In the scientific world, the quality and accuracy of the data are one of the biggest concerns that every scientific experiment has to consider. In the big data era, because the sources of the data are numerous and the qualifications of the organizations or individuals who collect the data are not equal, the quality and accuracy of the data are less controllable.
- **Value** refers to the usefulness of the information and knowledge we can derive from the data. Therefore, value is the most important V of Big Data. In any applications of big data, we have first to question what the value we can get from the big data.

Contrary to the traditional data management and analysis technologies, big data management and analysis must consider and properly deal with big data's five-V characteristics. Significant progresses have been made in both big data management, which deal with data capture, curation, archive, storage, cataloging, discovery, search, access, sharing, quality control, privacy, and etc., and big data analytics, which includes big data analysis, transformation, mining, visualization, knowledge discovery, and etc. However, challenges still exist in all above areas.

The purpose of big data analytics is to derive useful information and knowledge from big data while the main purpose of the big data management is to make big data analytics possible and feasible. In traditional data analysis, causal relationship among the variables is normally sought from data samples. Because of the number of variables, the volume of data, and the uncertainty in the data quality involved in the big data analytics, correlation or probability relationship are usually sought by analyzing whole datasets instead of samples.

3. AGRO-GEOINFORMATICS

Agro-geoinformation, the agricultural-related geo-information, is the key information in the agricultural decision-making and policy formulation process [4]. Agro-geoinformation is derived from agricultural-related geospatial data, or call agro-geodata. Agro-geoinformatics is a discipline that collects and manage agricultural-related geospatial data, derive agro-geoinformation and knowledge from the data, and applies the derived information and knowledge to solve management and decision-making issues in agriculture [4]. Therefore, agro-geoinformatics deals with the entire life cycle of agro-geoinformation transformation, ranging from data collection, information and knowledge derivation, and applications of information and knowledge in the agriculture domain. It includes both the agro-geodata management and agro-geodata analytics.

Agro-geodata is a kind of geospatial data, which is defined as the data that can be located on Earth (datasets with a spatial/location component). Studies have shown (or claimed) that more than 80% of data the world has collected is geospatial data [5]. Therefore, the geospatial data is big data. In fact, geospatial data have all five-V characteristics of big data.

4. EXAMPLES OF BIG DATA APPLICATION IN AGRO-GEOINFORMATICS

In the past several years, numerous big data management and analytic technologies have been developed. Many of them are general technologies that can be adopted by disciplinary big data applications. One of the most notable technologies to deal with big data is the cloud computing and associated Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [6]. Numerous software packages to deal with big data management and analytics, either on cloud or cluster platforms, have been developed. A lot of those packages are released as open-source software, freely available to all interested users. In the agro-geoinformatics domain, those general big data technologies have been adopted to deal with the big agro-geodata.

In addition to common five-V characteristics of big data, big agro-geodata also have their special features, particularly, the multi-dimensionality and spatial/temporal characteristics [7]. The agro-geodata may cover wide spatial area, up to the entire Earth. Because of the special features of agro-geodata, special big data management and analytics methods have to be developed. For example, both the spatial and temporal co-registration processing of multi-source and multi-modal data are essential for handling big agro-geodata.

In the past several years, the Center for Spatial Information Science and Systems at George Mason University has worked on a number of research projects, which deal with big agro-geodata for supporting the agriculture decision-making. Examples of such projects include:

- Development of geospatial sensor web technology to coordinately and purposely collect agro-geodata from in-situ, air-borne, and satellite-based imagery and non-imagery sensors [8]
- Web-service based near-real-time global agricultural drought monitoring systems [9]
- Web-service based near-real-time U.S. crop condition and progress monitoring system [10, 11]
- Seasonal agricultural drought prediction with data mining technology [12]

All of those projects have resulted in operational monitoring and decision support capabilities, which were not available before. The projects have adopted or adapted some general big data technologies and developed some domain-specific big agro-geodata technologies. The combination of adoption and development proves to be a successful approach to apply big data technology in the agro-geoinformatics discipline.

5. CONCLUSION

Agro-geoinformatics deals with collecting, managing, analyzing agricultural-related geospatial data, which are domain-specific big data. Through the adoption and adaptation process, the general big data technologies are very useful in agro-geoinformatics, but cannot solve all technology needs in dealing with big agro-geodata. Development of agro-specific big data technology is a necessary supplement to adoption of general big data technology. The combination of adoption of general big data technology and development of agro-specific big data technology prove to be a good strategy for applying big data technology in agro-geoinformatics.

11. REFERENCES

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers, Big data: The next frontier for innovation, competition, and productivity,*citeulike*:9341321, 2011.

[2] https://en.wikipedia.org/wiki/**Big_data**

[3] P. Hitzler and K. Janowicz, 2013. Linked Data, Big Data, and the 4th Paradigm. *Semantic Web*, *4*(3), pp.233-235.

[4] L. Di and Z. Yang, Foreword to the Special issue on Agro-Geoinformatics—The Applications of Geoinformatics in Agriculture. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 7(11), pp.4315-4316, 2014

[5] S. Hahmann, D. Burghardt, and B. Weber, "80% of All Information is Geospatially Referenced"??? Towards a Research Framework: Using the Semantic Web for (In) Validating this Famous Geo Assertion. In Proceedings of the 14th AGILE Conference on Geographic Information Science., 2011.

[6] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. A view of cloud computing. *Communications of the ACM*, *53*(4), pp.50-58, 2010

[7] L. Di, Distributed geospatial information services-architectures, standards, and research issues. *The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, *35*(Part 2), 2004.

[8] L. Di, February. Geospatial sensor web and self-adaptive Earth predictive systems (SEPS). In *Proceedings of the Earth Science Technology Office (ESTO)/Advanced Information System Technology (AIST) Sensor Web Principal Investigator (PI) Meeting, San Diego, USA* (pp. 1-4), 2007.

[9] M. Deng, L. Di, W. Han, A. L. Yagci, C. Peng, and G. Heo, Web-service-based monitoring and analysis of global agricultural drought. *Photogrammetric Engineering & Remote Sensing*, *79*(10), pp.929-943, 2013.

[10] Z. Yang, G. Yu, L. Di, B. Zhang, W. Han, and R. Mueller, Web service-based vegetation condition monitoring system-VegScape. In Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International (pp. 3638-3641). IEEE., 2013.

[11] W. Han, Z. Yang, L. Di and R. Mueller, CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Computers and Electronics in Agriculture, 84, pp.111-123, 2012.

[12] M. Deng, L. Di, G. Yu, A. Yagci, C. Peng, B. Zhang and D. Shen, Building an on-demand web service system for global agricultural drought monitoring and forecasting. In Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International (pp. 958-961). IEEE, 2012.