OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

A STUDY ON ENVIRO-COMPUTING: KEY CHALLENGES

Nishu Sethi[1], Neha Bhateja[2]

Assistant Professor, Amity University Haryana, Gurgaon, India[1 2]
sethinishu@gmail.com[1],bhateja.neha@gmail.com[2]

Abstract: **The most expressive and the basic research questions are integrated from various areas like mathematics, environmental science, computer science, biology etc. The new advanced development in computer science binds the gap with environmental science to tackle the data related to environmental models. In the environmental modeling area, various technical issues occur due to computational challenges like storage, retrieval, visualization, and processing. Our paper will give a review about these computational challenges in computing the environmental values.**

Keywords *– Computer science, Environment Science, computation, data analysis, big data, data assimilation.*

----------∴∴∴----------

I INTRODUCTION

Environmental science research is dependent on the analysis and grouping of large amount of information which is obtained through the observation sources and large scale deployments of sensors. The researchers are working on the global societal challenges which are collectively tied to the stability of our natural ecosystems. These challenges are fundamentally multifaceted and force the scientists to act as a team across penalizing precincts. Infrastructure supports the researchers to interact with different data sources and analytical tools, but still it cannot completely cover the entire ecosystem which supports the study of environment science [1,2]. Hence interoperability acts as a new challenge so that researchers can interact with all the research resources available to them, which allows to integrate and perform the innovative multifaceted research. This requires a realistic bridge between different scientific communities that can use a generalized framework for defining or which can relate similar concepts in different discipline. The formatter will need to create these components, incorporating the applicable criteria that follow.

II CHALLENGES IN ENVIRONMENTAL SCIENCE

1. Data Assimilation

For any environment problem, a number of factors or variables are responsible for it. These problems are generally stochastic problems rather than deterministic[5,6]. Data assimilation is one of the method to estimate these variables. It is an approach which works the combination of observation data and mathematical models.

The challenges of data assimilation are related to technical as well as general aspects in environmental science. The most significant challenges are:

a) The assimilation of the coupled data due to the coupling of the different elements of the environment for example, link between the atmosphere and the ocean, link between the atmosphere and land.

b) Assimilation of grouped data to represent for natural variability and the errors in the environmental system.

c) The main challenge of data assimilation is due to the integration and consolidation of the different modelling approaches [7].

2. Big Data, Earth Observation and applications

Environment and Earth sciences both are benefited through the techniques of Big Data Analytics which supports dispensation of huge amount datasets. Earth observation is produced from the simulations and observations [9,10]. Heterogenous data models, formats, geospatial information, complexity of processing the datasets are the specificities of Earth Science data and applications. Hence, customized techniques and tools are required for Big Earth Data Analysis.The big data has high impact on the observation of global earth system and on the digital infrastructure due to its 'V's' dimensionalities known as volume, variety, velocity, veracity and visualization.

Challenges of Big Data, Earth Observation and applications:

a) Multilevel analysis: Acquiring the data by ground survey techniques for multilevel analysis is very difficult, complex and time consuming.

b) Analysis methods: The joint classification of multistrategy takes benefit from complementarity of heterogeneous methods.

c) Multitemporal methods: This includes the opportunity for analysis of new type, where the Observations are irregular and clouds can hide the phenomenon also.

d) High frequency of observations: It becomes unrealistic to process the data for each new observation of high frequency big data.

e) Background knowledge: To classify the urban objects Some enterprises have focused on the use of domain knowledge. To build a knowledge base is tough task because knowledge is held by domain experts and is implicit which results in lack of formalization and exploitation of knowledge.

f) Lack of expertise: In what manner can we relate sufficient examples by class with high number of classes? This process is time consuming, labor intensive and requires that the experts have defined the types of required changes.

g) Scalability: This can be the major concern because the size of the data is not fixed and also it is large , it becomes difficult to exploit huge distributed dataset, which cannot fit into the memory.

h) Quality of data/knowledge: The raw data consists of errors, without the knowledge of data it becomes difficult to find, evaluate and correct errors or how to trust the expert ?

i) Robustness of algorithms : How to interpret errors in raw data (image) or segmentation data ,or incoherency in knowledge base.

3. Data analysis and challenges:

The main objective for the data analysis is to understand the relationships between features and to develop methods for mining which can accurately predict observations for the future. The complexity of data must be analyzed and scaled to process the data. Various devices are available which can generate increasing amount of data[9,10] .Hence the speed for mining the data(structured+ unstructured) has increased.

Techniques for data analysis:

a) Data Mining Algorithms: These can automatically discover useful patterns in the large dataset.

b) Cluster Analysis: These can sort different objects into groups so that degree of association between two objects is maximum, if belongs to same group else it is minimum.

c) Correlation Analysis: These measure the direction of relationships between two variables and the strengths of association between them. The value of correlation strength of variable lies in between +1 and -1.

d) Statistical Analysis: These measures gather, analyze, interpret, present, and organize the information.

e) Regression Analysis: This is the most common and basic technique that investigates the relationship between an independent and dependent variable.

Challenges:

a) Heterogeneity: Heterogonous data could be Binary, Continuous, categorical like-Images, trees, Graphs, Maps, ranking, Spatial Information etc .This exists because of excessive variations in data sets.

b) Scalability: This refers to the issues for Infrastructure Choices, Component Integration, Problem-Solving Strategies , Oversight, etc.

c) Accuracy: It is believed for most people that more the data being analysed more is its accuracy. But it involves certain challenges as all the data being relevant for the analysis may not be accurate, normalized, segmented or integrated.

d) Complexity: Complexity is related to as the space or time complexity which varies from data to data. The function describes the efficiency for the amount of data the it process [7].

e) Storing/Sharing/Publishing: These are the 3 pillars that exist in the life cycle of big data. The available infrastructures is used to manage and publish data which is designed to fulfill the requirements of project or domain specific., that results in the repetitive expansion of registered solutions.

f) Security: The wide range of data sets includes: unstructured data and binary data like videos, audios and images that offers vital intelligence in business risks which are far-off the Information Technology. The unknown security threats still concealed with the use of advanced statistical modeling and predictive analysis.

Retrieve/Reuse/Discover: Data quality, data preservation and value addition are safeguard by retrieving the data which can reuse the existing data to acquire new and valuable information. The reusability defines the meaning of the published data.

III CONCLUSION

In the environmental science, different elements are linked with each other like atmosphere and land, the atmosphere and the oceans etc. Due to the integration of multifaceted concepts in the environment, the representation of data becomes difficult. The advanced technologies of computing helps to fill this gap in the environmental science using concept of big data , data analysis tool etc.

REFERENCES

[1] Martin, Paul, Chen, Yin, Hardisty, Alex, Jeffery, Keith and Zhao, Zhiming 2017.Computational challenges in global environmental research infrastructures. In: Chabbi, Abad and Loescher, Hank W eds. Terrestrial Ecosystem Research Infrastructures: Challenges and Opportunities, Boca Raton, FL. USA: CRC Press, Taylor & Francis Group, pp. 305-340

[2] Kim. H., Seungdo Kim, and Bruce E. Dale. 2009. Biofuels, Land Use Change, and Greenhouse Gas Emissions: Some Unexplored Variables. Environmental Science and Technology, 43 (3): 961–967.

[3] UNEP. 2007. Global environment outlook 4 (GEO4). Technical report, United Nations Environment Programme (UNEP).
[4] Garey MR, Johnson DS. Computers and Intractability: A Guide to the Theory of NP-Completeness. New York: W. H. Freeman; 1979
[5] Clark, J S, and Gelfand, A E (2006), "A Future for Models and Data in Environmental Science", TRENDS in Ecology and Evolution, 21(7), pp 375-380.
[6] Hall, J, and O'Connell, P E (2007), "Earth Systems Engineering: Turning Vision into Action", Civil Engineering, Proc Institution of Civil Engineers, 160(CE3), pp 114-122.
[7] Little, R G (2002), "Controlling Cascading Failure: Understanding the Vulnerabilities of Interconnected Infrastructures", J Urban Technology, 9(1), pp 109-123.
[8] Lin, Z, and Beck, M B (2007b), "Understanding Complex Environmental Systems: A Dual Thrust", Environmetrics, 18(1), pp 11-26.
[9] Nishu Sethi," A Comprehensive Study on Data Science and its Practical Applications", International Journal on Recent and innovation trends in computing and communication. Issue 5, Volume 5, pp 573-575.
[10] Neha Bhateja"A Business approach to Data Driven Science- A study", International Journal of Advanced Research in Computer Science, Volume 8, Issue 5, pp 2483-2485.

SSWM-17
National Conference on Sustainable Solid Waste Management

Organized by:

Department of Civil Engineering,
Amity School of Engineering & Technology
Amity University Haryana