Potential of Integrated Data Geography as a new Systems Engineering Approach

Dr. Tapobrata Lahiri (tlahiri@iiita.ac.in)
Mr. Rajnish Kumar, Mr. Manoj Kumar Pal and Mr. Nilotpal Pramanik
Indian Institute of Information Technology, Allahabad, Prayagraj, UP, India

This article introduces the concept of "Integrated Data Geography (IDG)" approach which stems from the concept of Integrated Geography or Integrative Geography (IG). IG in its definition is the branch of geography that describes and explains the spatial aspects of interactions between human individuals or societies and their natural environment, called coupled human–environment systems [1-2]. Basically geography is of two types, spatial (i.e., environmental geography), and, human geography. The third advent is integrated geography which is of course assimilation or coupling of these two information to gain a holistic or connectionist view of the system as a "integrated system". For example, social and religious practices of a human group at a particular geographical location might have been rooted from the need to withstand geo- environmental pressure within that particular geographical location. Therefore, if we find the same practice prevalent in some very distant geographical location, then possibility of migration of same human race may be mooted upon.

Continuing with the concept of IG, we thought that it might also be interesting to employ this approach as an additional component of Systems Engineering to get more in-depth information about the system through pulling in already reported (i.e., results of publications) pieces of information and putting them together under a framework of Data Science, whereas, Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data [3]. It employs techniques and theories drawn from many fields within the broad areas of mathematics, statistics, information science, and computer science, in particular from the sub-domains of machine learning, classification, cluster analysis, uncertainty quantification, computational science, data mining, databases, and visualization.

In the backdrop of a huge amount of research reports being cumulatively piled up daily as publications, it appears to be necessary to tap this huge information treasure for effective analysis of a system from a Systems Engineering View. Proposal of this approach is distinctly different from Meta Analysis approach which deals with statistical summarization of same System Property or System Issue worked upon and published by various researchers. The systems engineering approach being proposed here stands on the premise that, for a particular system, its different issues or properties worked upon and published by different group of researchers may be judiciously combined to achieve a particular new Systems Engineering goal about this system without doing any further experiment or work on this system. As already stated, under current circumstances, a substantial amount of reports of already performed research on different subsystems of a particular system appears to make such approach plausible. For further clarification on this proposed approach, following examples may be found to be useful.

First example is about phishing in biometrics based personal identification. For attempt of phishing in a fingerprint based personal identification system, the target may be to detect data-defect for a typical personal fingerprint. In this regard, it may fail for a particular type of derived data or feature, but may be successful in detecting phishing attempt through use of different features and combining different feature-outputs under rules of assimilations.

Second example is a VLSI design which is waiting to be optimized as target hardware or diagnosed after a fault. For both of these cases, the standard practice is to achieve it through trials and not under any established systems engineering framework or anything similar to IDG.

Third example is an automobile system which is to be optimized for the purpose of the intended users (say Defense) or diagnosed after a defect. For this purpose also, no data science or systems engineering framework is practiced.

Fourth example is of medical diagnostic or therapeutic system, where it is of frequent need to assimilate information obtained from various tests towards diagnosis and subsequent planning of treatment.

To give a mathematical framework of this concept, it may be said:

For a pool of logical lemmas,

L = {Li}, for i = 1, 2, ..., N,

if,

i) all of these logical lemmas serve as supports to prove our proposition to build a theorem, T = f(L) for f is a function in broader sense which is derived through judiciously made inter-connectivity of all the members of L, and also,
ii) these lemmas represent complimentary form of necessary and sufficient proof of T,

we may consider the analysis of the system is compatible and doable under the paradigm of IDG. From the above-representation although it appears to be very obvious and commonly practiced method of analysis, which is actually so among crime-investigators or detectives, use of logical lemmas in Science and Engineering is yet to be formally introduced.

The following argument appears to further strengthen the basis of such new Systems Engineering component along with its difference from the existing Data Science approaches. One such example is, Meta Analysis, which is a statistical analysis possible under a condition that all different published results should have been produced under nearly same methodological steps only so that a statistical estimate of an intended goal parameter may be obtained. However, in IDG it is not mere estimation of a parameter, but to obtain necessary and sufficient support from outcomes (i.e., Logical Lemmas in technical sense) of analysis of different sub-data's extracted from different sub-systems of the whole system to assert about the proposed behaviours of the system. To say geographically, these sub-datas can be thought of representatives of different entities (like human) attached to different geographical part of the data derived from different sub-systems. Since the data in this regard may be even a typical nominal scale data (e.g., a class decision), therefore, the reach of this approach is supposed to be beyond the scope of usual Systems Engineering approach. The need of such IDG approach is increasing day by day due to availability of results of huge amount of experimentation being carried out over a system.

Reference

1. G., Moseley, William (2014-01-01). An introduction to human-environment geography : local dynamics and global processes. Wiley-Blackwell. ISBN 9781405189316. OCLC 921583361.

2. Moseley, William G.; Perramond, Eric; Hapke, Holly M.; Laris, Paul (2014). An Introduction to Human-Environment Geography. Wiley Blackwell. pp. 26–27.

3. Hayashi, Chikio (1998-01-01). "What is Data Science? Fundamental Concepts and a Heuristic Example". In Hayashi, Chikio; Yajima, Keiji; Bock, Hans-Hermann; Ohsumi, Noboru; Tanaka, Yutaka; Baba, Yasumasa. Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Japan. pp. 40–51. doi:10.1007/978-4-431-65950-1_3. ISBN 9784431702085.

About the authors

Dr. Tapobrata Lahiri Obtained his Ph.D. Degree from Kalyani University, Calcutta, India. He is currently serving as Professor and Dean (Academic and Research), Indian Institute of Information Technology-Allahanad, India. He has been also selected as Research Advisor by Nan Yang Academy of Science, Singapore in 2018. His main field of research includes Application of Numerical Analysis Techniques (Machine Learning, Artificial Intelligence, Optimization, Systems Modelling and Simulation, Fractal Dimensional Analysis) to improve Prediction and Automatic Expert/Diagnostic Models for Biomedical application.

Rajnish Kumar received his Masters in Information Technology (Specialization in Bioinformatics) from Indian Institute of Information Technology in 2012 and submitted his PhD in 2019 in the Department of Applied Sciences under the supervision of Prof. Tapobrata Lahiri. Currently he is working as Bioinformatics Scientist in Neuberg Supratech Laboratory, Ahmedabad, Gujarat, India. His area of interest includes Analysis of genomics and transcriptomics Next generation sequencing data, Bioinforamtics Algorithms and Machine Learning.

Manoj Kumar Pal received his Masters in Software Engineering from Motilal Nehru National Institute of Technology in 2015 and currently pursuing his PhD in the Department of Applied Science under the supervision of Prof. Tapobrata Lahiri. His area of interest includes of machine learning, digital image processing, soft-computing, and optimization techniques.

Nilotpal Pramanik is a fourth Year student of Dual Degree Bachelor's and Master's in Information Technology (Specialization in Robotics at Indian Institute of Information Technology, Allahabad, India. His area of interest includes Machine Learning and Computational Modelling.