ISBN: 978-1-948012-16-4

Contemporary Issues in Computing (CIC)

DOI: <http://doi.org/10.26480/cic.01.2020.118.120>

CrossMark

OPTIMIZING THE VALUE OF BIG DATA: ROLE OF COMPUTATIONAL SOCIAL SCIENCE

Renji George Amballoor* and Shankar B. Naik

Directorate of Higher Education, Govt. of Goa, Porvorim, Goa 403001, India

*Corresponding Author e-mail: amballoor@gmail.com

This is an open access article distributed under the Creative Commons Attribution License CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 26 October 2020
Accepted 27 November 2020
Available online 03 December 2020

ABSTRACT

Computational Social Science is an emerging discipline which provides better opportunity to collect, analyze and interpret big data sets on different aspects of human behaviour. The quantum expansion of digital technologies and the dependence on them in everyday life has generated huge diverse and innovative big data sets on social phenomena which otherwise was beyond the reach and expectation of social scientists. The new sources of data, type, quantity, etc. required an interdisciplinary approach for optimizing the data usage. In the new discipline, the research from collaborating subject domains employ agent based models to derive stories from big data. The agent based computational social science has ensured collaboration between social scientists and computer scientists for tapping and for drawing meaningful conclusions from the big data created through the use of digital technologies.

KEYWORDS

Big Data, Computational Social Science, Agent Based Models, Social Behaviour, Social Science, Computer Science.

1. INTRODUCTION

The percolation and convergence of cyber-physical systems (CPS) (Gill, 2006) and digital twins (DTs) (Grieves, 2014) along with the Internet of Things (IoT) architecture has witnessed the amoebic expansion of big data footfalls. Convergence of digital technologies contributed towards the creation, collection and processing of huge amount of big data from network of people, things, machines, etc (Patel et al. 2016). The process in enabled with smart devices and sensors which continuously collect and process data on real time basis.

Today data is being generated from satellite images, medical devices, sensors, network of people and machines, digital media, social and technological stimulations, etc. The big data so generated differ in its complexity, size, granularity, reliability, scope, etc. According to National Science Foundation (2012), huge, distinct, longitudinal and distributed datasets created through digital transactions, emails, videos, geo-locations, mobile towers, sensors, etc. often referred to as digital exhaust or digital breadcrumbs for drawing inferences on individual and collective social behaviours. The huge size, complexity, high dosage of data availability has made the management and analyzes of big data a complex task. According to Gartner Glossary, Big data is high volume, velocity and variety information asset. Chen (2014) is of the view that big data consists of data sets containing information on socio-economic trends and human behavior. The situation demands new methods & technologies for data driven discovery and decision making.

Computational Social Science is an emerging domain wherein the Social

Scientists using data processing and data science computation tools (Python, Matlab, Scilab, etc) to analyze data about social phenomena people their relationships, interactions, behaviours, actions and counter reactions. According to Holthoefer et al. (2016) Computational Social Science provides abstracted models and methods, large storage, algorithms, computational power, social hypotheses and conceptual framework for decoding the results. Computational Social Science provide scientific scenarios for addressing social complexities like multipronged interconnections, interdependence, interventions, leapfrogging dispersions, etc. for different magnitude, layers and difficulty levels.

Many collective phenomena like socio-political, cultural, technological and demographic changes witnessed in the economy and society are not because of one individual agent's behaviour but because of the complex interactions among various heterogeneous agents in diverse environments over a period of time. According to Anderson (1972), social systems exhibit emergent behaviour i.e. the behaviour of entities like group, crowd, etc. at one scale of reality is not traced to the characteristics of the entities at the scale below. The data generated by the changes were huge, dynamic and complex requiring modelling.

Watts (2016) is of the opinion that the existing theories of social behaviour and change are inadequate to give a realistic picture of social phenomena which is complex, multi-scale and exhibits emergent nature. The magnitude of the problem multiplies when it is accepted that traditionally it has been difficult to collect observational data about some social phenomena on such a large scale. Further the researchers are also

Quick Response Code	This paper was presented at	Access this article online
International Conference on Contemporary Issues in Computing (ICCIC-2020) - Virtual IETE Sector V, Salt Lake, Kolkata From 25th-26th July 2020	Website: www.intelcomp-design.com	DOI: 10.26480/cic.01.2020.118.120

handicapped at inferring the cause and effect from such data. The situation becomes more complex due to the difficulty experimentally fiddling with the large number of entities in a social problem. These constraints which are at present hindering the growth of social sciences can be addressed by having an alliance with computational sciences which has resulted in the new paradigm of computational social sciences (Lazer et al. 2009).

In 2009 a battalion of 15 researchers published a paper in Science proclaiming the birth of a new discipline computational Social Science (Mann, 2016). According to Lazer et al. (2009) CSS is integration between social and computational sciences which will use observational data, experiments on virtual laboratories and computational modelling.

Wallach (2016) remarked that researchers are at threshold of entering a new world of computational social science. For studying the social behaviour, this new breed of social scientists are using computational methods (Alvarez, 2016, Trilling, 2017). The armory of computational social science includes the following: (1) huge complex data sets (2) digital traces and other naturally occurring data (3) uses algorithms to understand the data sets (4) study human behaviour in different social science contexts. In computational method, the researchers with big data and algorithms can go beyond mere escalating of previous research (Atteveldt, 2018).

The equation based models (EBM) like econometric and analytical models captures only the macro behavior of the system without capturing the triggers for the behavior of individuals (Singh 2019). This was the dominant modeling tool used in various social sciences including Economics for long time. But EBM fails to explain how the behaviour of individual agents like consumers will influence the macro behaviour of a system. Further the power of the model to capture the emergent behaviour was very poor (Sabzian 2019).

The agent based models (ABM) originated by Enrico Fermi in 1930s to address the problem associated with transporting neutrons through matter. During 1990s the model found its application into different disciplines like Biology and Ecology (individual based models) physical sciences (Monte Carlo simulations), Computer Science (multi agent systems), Economics (Agent based computational economics models), Defence planning, Epidemiology, Social Sciences, Strategic studies, etc. Agent based computational economics models illustrate the behaviour of a system (economy) by simulating the behaviour of each individual agent (consumers) in it. The model adopts a bottom-up approach i.e. how individual decisions of consumers can be grouped together to understand the complex behaviour observed in the real economy.

The agent based computational social science provides researchers the proficiency to collect analyze data of huge depth, breadth and scale on human behaviour and interactions. This method allows the researcher to locate the data that meet the pre-planned experimental conditions without taking the entities into laboratory framework for conducting experiments. The agent-based model is of vital importance in computational social science, because social theories revolve around people. Social science deals with different aspects of human beings and its bandwagon includes economics, sociology, political science, psychology, anthropology, etc. The thoughts, beliefs, likings, preferences, values, desires, actions, reactions along with the different processes involved create social complexity.

The agent based model which are qualitative in nature can give inputs on what could occur as an emergent situation rather than to predict the exact nature of event that will occur (for example, what are the scenarios that can occur in future due to the slowing down of the manufacturing sector) (Turrell, 2016). The agent based models while studying the interactions of many heterogeneous agents to understand the overall behaviour of the economy results in the creation of huge amount of data. Simulation typically generates huge amounts of data (Axelrod, 2006). In order to capture the data, analyze and draw patterns, computational and mathematical modeling is required for social scientists and economists. Unless the disciplines come out from their silos, their mutual power in addressing the concerns of the economy and society will be minimum.

Figure 1: An Agent Based Model

Figure 1 shows how the interactions between heterogeneous agents and actors in a social phenomenon can be captured and analyzed using the agent based models where in the agents are internet users having their profiles created on some social website or messenger app. Each user in this case observes a phenomenon or an event such as a cricket match, natural disaster, celebration, accident, political issue, etc and as a reaction to the stimulus from these events, posts messages on the social websites/messenger apps. The receivers of these messages accordingly respond to them. These messages and responses reflect the interaction between the agents. Different users can have accounts on different social websites/messenger apps thus making the agents heterogeneous.

These interactions in the form of messages can be mined to find patterns containing frequent words. These frequent words would be the ones trending high on social media thus reflecting the topics relevant at that time and place. In this example the behaviours (messages) of individual online users are used to discover the overall trending topics in society at different levels of granularities of space such as city, state, country, and world and also at different levels of granularities of time such as trends in a day, week, month, season and year.

Figure 2 represents a real world situation in which the machine learning algorithms can input different layers of complex behaviour of agents in the society to get meaningful patterns. The input to the algorithm are values of attributes of an online shopping website user. The attributes could be gender, age and region. The algorithm combines this data with additional data such as current time and knowledge rules to predict the category of items along with the probability that the user will search for the items in that category. These patterns of knowledge can help online shopping companies to predict the potential list of items the user would be interested in, thereby helping the user in searching for the items of his/her choice and also increasing the chance of items being purchased by the user.

Figure 2: Machine Learning Algorithm with Multiple Output Scenarios

According to Anzola (2018), the agent based computational social science has four distinct features:

1. It is inter-disciplinary in nature due to the diversity in partnering research disciplines
2. High degree of dependence on technological resources and tools for collaborative work.
3. Pre-dominance of practical orientation over theoretical and philosophical concerns.

4. Since the discipline is based on agent based model, methodological homogeneity is present. The methodological framework is drawn from complexity science, computer science and social science.

Once the big data goes through the methodological framework of agent based computational social science, future scenarios and patterns can be deducted. If there were no alliance between social sciences and computer science, social scientists would have been handicapped to use big data sets for their research. Similarly, if the computer scientists were to work in watertight silos, the patterns emerging from big data would have just remained as mere patterns. The real meanings hidden behind these patterns can be visualized and understood only with the help of social scientists. Once the meaning behind the patterns is understood, the same can be used by policy makers for formulating policies and nudges which will benefit the society and their agents.

Many researchers have joined the bandwagon of agent based computational social science however; the numbers are not very impressive. The lack of popularity is basically due to the new framework made possible by dig data. Similarly, the computer scientists are busy writing new algorithms of understanding new patterns from big data. But they fail to understand the importance of social scientists for drawing meaning from these patterns. On the other hand, the social scientists have to collaborate with computer scientists for developing models and algorithms for capturing, mining and analyzing big data. According to Brynjolfsson & McAfee (2011), insights for outlining processes can be best possible when the researchers combine data analysis using AI along with intuitions of the people.

Once the common language of communication between social and computer scientists are established, the discipline of computational social science will skyrocket. This will also help the researchers to optimize the use of big data for the benefit of the society and its agents.

FUNDING

This research is neither sponsored nor funded by any agency, department or institution.

REFERENCES

- Alvarez, R. M. (Ed.). 2016. Computational social science: Discovery and prediction. Cambridge, UK: Cambridge University Press.
- Anderson, P. W. 1972. More is different. *Science* 177(4047):393-96.
- Anzola David. 2018. Knowledge Transfer in Agent based Computational Social Science. *Studies in History & Philosophy of Science*.
- AtteveltdtWouter van & Tai-Quan Peng. 2018. When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science. *Communication Methods & Measures*. Vol. 12, No. 23.
- Axelrod Robert. 2006. Advancing the Art of Simulation in the Social Sciences in *Handbook of Research on Nature Inspired Computing for Economy and Management*, Jean-Philippe Rennard (Ed.). Hersey, PA: IGI Global.
- Brynjolfsson, Erik & McAfee, Andrew 2011. *Race against The Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*. Lexington, Massachusetts: Digital Frontier Press.
- Chen Michelle 2014. Is Big Data Actually Reinforcing Social Inequalities? Accessed from <https://www.thenation.com/article/archive/big-data-actually-reinforcingsocial-inequalities/>
- Gill, H. 2006. NSF perspective and status on cyber-physical systems. Austin.
- Grieves, M. 2014. Digital Twin: Manufacturing Excellence Through Virtual Factory Replication (White Paper). Michael Grieves (University of Michigan) LLC (<http://innovate.ttu.edu/plm/documents/docmgr/912/1411.0DigitalTwinWhitePaperDrGrieves.pdf>).
- Holthoefer -Javier Borge, Yamir Moreno & TahaYasseri. 2016. At the Crossroads: Lessons and Challenges in Computational Social Science. *Frontiers in Physics* 4(37).
- Internet: <http://varma.ece.cmu.edu/CPS/Presentations/gill.pdf> [zulezttaufgesucht am 1.04. 2015].
- Lazer, D., A. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, and M. Gutmann. 2009. Social Science: Computational Social Science." *Science* 323(5915):721.
- Mann Adam. 2016. Computational Social Science. *Proceedings of National Academy of Science of USA (PNAS)*, 13(3).
- National Science Foundation, Solicitation 12-499: Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA), 2012, <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf>.
- Patel Keyur K et.al. 2016. Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges. *International Journal of Engineering Science and Computing*.
- Sabzian Hossein 2019. Theories and Practice of Agent based Modeling: Some practical Implications for Economic Planners. Accessed from <https://arxiv.org/pdf/1901.08932.pdf>
- Singh Kartik 2019. Introduction to Agent-Based Modelling. Agent-Based Modelling, Data Science. Accessed from <https://dimensionless.in/introductionto-agent-based-modelling/>
- Trilling, D. 2017. Big data, analysis of. In *The J. Matthes, C. S. Davis, R. F. Potter (Eds.), international encyclopedia of communication research methods*. New York, NY: Wiley Online Library.
- Turrell Arthur. 2016. Agent-based models: Understanding the Economy from the Bottom up. *Quarterly Bulletin - Q4*. Accessed from <http://www2.econ.iastate.edu/tesfatsi/ABMOverview.BankOfEngland.ATurrell2017.pdf>
- Wallach, H. 2016. Computational social science: Towards a collaborative future. In R. M. Alvarez (Ed.), *Computational social science: Discovery and prediction* (p. 307). Cambridge, UK: Cambridge University Press.
- Watts Duncan J. 2016. Computational Social Science: Exciting Progress and Grand Challenges. Microsoft Research. Accessed from <https://www.nap.edu/read/18558/chapter/6>.