Mini Symposium on Big Data and Large Scale Computing

December 27, 2017

125th Birth Anniversary of

Prof. P.C. Mahalanobis celebration

Organized by

Indian Statistical Institute,

Delhi Centre

PC Mahalanobis 125

Contents

Welcome to SBGC17

On behalf of the organizers of SBGC17, I welcome you in the mini Symposium on big data and large scale computing will be organized on December 27, 2017 at Indian Statistical Institute, Delhi Centre as a part of 125th Birth Anniversary of Prof. P.C. Mahalanobis celebration. In today's world, ability to store, manage, and analyze large-scale data generated by humans and machines has a critical impact on decision making in every business, scientific discovery, social and environmental challenges. This mini symposium deals with different aspects of data science, efficient data processing, large scale computing (distributed and parallel computing).

The objective of this symposium is to provide a forum for new developments and applications in the area of big data and large scale computing. Leading scientists, experienced researchers and practitioners, as well as younger researchers will come together to exchange knowledge in this area of big data-statistical data analysis &computing and to explore new directions in large scale data analysis.

The symposium topics include (but not limited to):

1.Big Data Discovery and Its Impact on Decision Making

2. Big Data Search Architectures, Scalability and Efficiency

3. Data Acquisition, Integration and data mining

4. Computational Modeling and Data Integration

5. Algorithms and Systems for Big Data Search

6. Privacy Preserving in Big Data Collection

7. Methodologies for large scale computing

8. Case Studies

S. K. Neogy
Organizing Committee Chair

Committees

Organizing Committee

S. K. Neogy (Chair), Dipti Dubey, R.B. Bapat, Arunava Sen, Prabal Roy Chowdhury

Programme Co-ordinating Committee

Dipti Dubey, R. Chakraborty and Praveen Pandey

Facilities Committee

R. Chakraborty, R. C. Satija, Simmi Marwah, Praveen Pandey, Sujan Dutta, Parama Gogoi and Srinivas

Mini Symposium on Big Data and Large Scale Computing

Program Overview

Session Details

Registration: 9:30

January 27, 2017 Time: 10:00 -10:15 Venue: Auditorium

Opening Remarks, About symposium, Vote of Thanks

Time: 10:15 -12:00 Venue: Auditorium

Session I

1.	**10:15-11:10- Rajeeva Karandikar** (Chennai Mathematical Institute) Role of Statistics in the Era of BIGDATA
2.	**11:10-12:00- Krithi Ramamritham** (Department of Computer Science and Engg.(CSE) IIT Bombay) The Role of Analytics in Smart Energy

Tea Break: 12:00 -12:20

Session II

December 27, 2017 Time: 12:20 -13:10 Venue: Auditorium

1.	**12:20-13:10- R. K. Agrawal, (**School of Computer & Systems Science, Jawaharlal Nehru University, New Delhi) A Visit to Artificial Neural Network to Deep Learning

Lunch Break: 13:10 -14:00

December 27, 2017 Time: 14:00 -15:30 Venue: Auditorium

Session III

1.	**14:00-14:45- Dinesh Garg,** (Computer Science and Engineering, IIT Gandhinagar, India) Latent Space Embedding Models for Information Retrieval in Question-Answer Corpus**.**
2.	**14:45-15:30- S K Neogy** (Indian Statistical Institute, Delhi Centre) Statistical Methods for Analyzing Big Data: A Research Paradigm Shift

Tea Break: 15:30 -15:45

Session IV

1.	**15:45-16:15- B. Majumdar** (visiting faculty, Indian Statistical Institute , Delhi Centre) Role of Big Data in Time Series Analysis
2.	**16:15-17:00- Santosh Singh** (Shiv Nadar University): Analysis of Big data in Medical Imaging
3.	**17:00-17:15- Sanjay Prakash Bhagat** (Ministry of New and Renewable Energy, Government of India): Role of Big Data in the Development of Renewable Energy in India
3.	**17:15-17:30**: **Panel Discussions** (S K Neogy, Santosh Singh, Dinesh Garg)

High Tea: 17:30 -17:45

ABSTRACT OF THE TALKS

Role of Statistics in the Era of BIGDATA

Rajeeva Karandikar,

Chennai Mathematical Institute
H1 Sipcot IT Park, Siruseri, Kelambakkam 603103, Chennai, India

Abstract

Statistics evolved as a science in an era where the amount of data available was small and efforts were onto extract maximum information out of that.

Are the techniques developed in that era relevant anymore in the era of BIGDATA?

We will discuss this. We will illustrate via examples that while some techniques may no longer be relevant, the concepts are as relevant in this era as they were then.

The Role of Analytics in Smart Energy

Krithi Ramamritham

Department of Computer Science and Engg. (CSE) IIT Bombay

Smart grids have been heralded as the key enabler of cleaner, cheaper and more reliable energy. They incorporate diverse energy sources, advanced monitoring, demand-side management and the ability to "self heal". The success of smart grids lies in the development of effective solutions for a) Demand-supply management incorporating intermittent, renewable, energy sources; b) Monitoring and sensing to understand energy generation and consumption patterns; and c) Distributed information management and control strategies. The talk will cover these topics and show how timely data analytics plays a crucial role in addressing these problems.

A Visit to Artificial Neural Network to Deep Learning

R. K. Agrawal,

School of Computer & Systems Science,

Jawaharlal Nehru University,

New Delhi.

Artificial neural network, a supervised learning approach, is one of the well-known classification method in machine learning. A decision model is build based on available training samples. The decision model is further used to predict label corresponding to a new sample. However, the training samples need to be represented in terms of relevant features to build decision model. To extract relevant features from a given data to achieve maximum performance of the decision system is not trivial. One has to explore a large set of feature extraction methods to obtain relevant features, which is challenging and difficult.

Recently motivated by the structure and function of the brain, Deep learning method is proposed. It is a subfield of machine learning which aims at automatically learning features hierarchies to represent high level abstractions from the large volume of labeled data. In past few years, deep learning is successfully applied to many domains such as Natural language processing, Speech processing, Computer vision and Image processing.

In my talk, I will discuss basics of Artificial neural network and deep learning architecture along with its implementation. Some experimental results in this direction will also be discussed.

Latent Space Embedding Models for Information Retrieval in Question-Answer Corpus

Dinesh Garg,
Computer Science and Engineering,
IIT Gandhinagar, India

Community-driven Question Answering (CQA) systems such as Yahoo! Answers have become valuable sources of reusable information. CQA retrieval enables usage of historical CQA archives to solve new questions posed by users. This task has received much recent attention, with methods building upon literature from translation models, topic models, and deep learning. In this talk, I will present a CQA retrieval technique that embeds question-answer pairs within a unified latent space preserving the local neighborhood structure of question and answer spaces. The idea is that such a space mirrors semantic similarity among questions as well as answers, thereby enabling high quality retrieval. Empirical analysis on various real-world QA datasets suggests effectiveness of our technique over state-of-the-art methods.

Statistical Methods for Analyzing Big Data: A Research Paradigm Shift

S. K. Neogy
Indian Statistical Institute Delhi Centre

Big data technologies and the corresponding fundamental research have become a research focus in academia. The emergence of big data has provided us unprecedented large-scale samples and computational problems, although we now have to face far more complex data objects. This takes big data as its research object and aims at generalizing the extraction of knowledge from data. Extracting meaningful information from a large data set may require methodological and interdisciplinary flexibility to create business intelligence for decision support. It spans across many disciplines, including information science, mathematics, social science, network science, system science, psychology, and economics. A number of challenges have been identified that need to be addressed while using Big Data for different applications.

Role of Big Data in Time Series Analysis

B. Majumdar

Guest Faculty

Indian Statistical Institute Delhi Centre

Time Series analysis and Forecasting is of great importance in various practical domains. It is an integral component for management, planning and decision making. Historically Statistical analysis of Time series data originates from the works of Yule and Kendal (1927). Since then the subject has grown as a full fledged area of research and applications. After the emergence of Big Data, large amounts of time series data are available from many heterogeneous data sources and varied application areas. The volume, variety and speed of generation of such data, introduced new directions to Statistical analysis of Time Series Data and forecasting. In this presentation, the role, scope and opportunities of Big Data in Time Series Analysis is explored. Also, the applicability of Big Data vis-a-vis the traditional approaches is considered.

Analysis of Bigdata in Medical Imaging

Santosh Singh

Shiv Nadar University

Role of Big Data in the Development of Renewable Energy in India

Sanjay Prakash Bhagat

Scientist C,

Ministry of New and Renewable Energy,

Government of India,

New Delhi, India

bhagatsp23@gmail.com

Big data analytics starts playing an important role in all areas, power sector is not the exception. Power industry is observing a transition to renewable energy after larger focus on renewable energy in last few years. Government of India has set a target of 175 GW of renewable energy installation by 2022. Renewable Energy is the intermittent power source. The integration of renewal energy sources like solar or wind with the grid is a challenge as its intermittency puts a question mark on the stability of the grid. In this paper the new data collection technique for utilizing and managing the renewable energy and its integration with the electrical grid are proposed. Paper also highlights on areas like energy monitoring, fault management, MIS report generation by using big data analysis tools like Hadoop, R and Python.

Keywords- Big data analysis, application in renewable energy, Grid Integration, Renewable Energy Management Centre(REMC).