# CALL for PARTICIPATION

## In conjunction with PREMI 2013

## IUPRAI Workshop on

# Big Data: A Soft Computing Perspective

**Date: 14.12.2013**

**Place: ISI, Kolkata**

The main challenges in handling Big Data lie not only in the four V's, namely, Volume, Variety, Velocity and Veracity, but also its Variability in content and structure which calls for new approaches for interpretation and understanding of the data. Big Data demands a revolutionary change in research methodology and in tools to be employed. Traditional Computational Intelligence techniques need a complete re-haul to handle these challenges. Soft Computing techniques can play a significantly important role in this due to their inherent capabilities of dealing with imprecision and uncertainty. This workshop aims to bring together researchers and practitioners who are working on applications of soft computing principles to handle big data challenges. The aim is to share the recent advances in research and also discuss the challenges and opportunities in the application of Soft Computing techniques to the emerging field of Big Data.

## Programme

**Morning Session**

| | |
|---|---|
| **11:30 A.M – 13:15 P.M** | **Session 1: Techniques & Challenges** |
| 11:30 A.M – 11-45 A.M | Opening Remarks: Prof. Sankar Pal |
| 11:45 A.M - 12-15 A.M | Recursive and Iterative clustering in granular hierarchical, network, and temporal datasets: Prof. Pawan Lingeras, Saint Mary's University, Halifax |
| 12-15 P.M - 12-45 P.M | Toward Granular Scalability of Analytical Data Processing: Prof. Dominik Slezak, University of Warsaw & Infobright |
| 12-45 P.M – 13-15 P.M | Ocius, minutus, exactius – the Olympic Theme for Next Generation Machine Learning: Prof. Jayadeva, IIT Delhi |

**Afternoon Sessions**

| | |
|---|---|
| **14:30 P.M – 15:30 P.M** | **Session 2: Applications** |
| 14:30 P.M – 15:00 P.M | Soft Computing Challenges in Algorithmic Trading: Saugata Mukerji, Marketopper Securities Pvt. Ltd. |
| 15:00 P.M – 15-30 P.M | Persona Based Mobile Marketing: Prateek Kapadia, Flytxt Mobile Solutions |
| **16:00 P.M – 17:00 P.M** | **Session 3: Technologies** |
| 16:00 P.M - 16-30 P.M | Mining large document repositories: Girish Keshav Palshikar, Tata Research Development and Design Centre, Pune, India. |
| 16:30 P.M – 17-00 P.M | Cognitive Computing: Raghavendra Singh, IBM Research |

## SPONSOR

## ORGANISERS

**Advisor: Prof. Sankar Pal**

**Co-ordinators: 1. Dominic Slezak 2. Santanu Chaudhury 3. Lipika Dey**

**Industry Interface: 1. Prateek Kapadia, Flytxt,**

## REGISTRATION

**Complimentary with registration of PREMI 2013 being held at ISI, KOLKATA**

**http://www.isical.ac.in/~premi13/**

**Registration fees for workshop ONLY: Rs. 3000.00.**

**Please visit PReMI 2013 Registration Page for details.**

# Workshop on Big Data: A Soft Computing Perspective

## DETAILS OF TALKS

A. **Title: Recursive and iterative clustering in granular hierarchical, network, and temporal datasets:**

   **Speaker: Pawan Lingeras,**

ABSTRACT

Clustering is one of the frequently used unsupervised data mining techniques for grouping similar objects. The proposed research program will investigate a novel iterative approach to clustering in a granular environment. An information granule represents an object. For example, a customer with certain purchasing patterns could be represented by an information granule. A granule is usually connected to other granules. For example, in a hierarchical environment, a customer granule will be connected to a number of product granules and vice versa. In a granular network, phone users are connected to other phone users. In a granular temporal environment, a daily pattern of events is connected to historical and future daily patterns. Traditionally, clustering of granules is done in isolation without any information on clustering of the connected granules. The primary theme of the proposed research is to simultaneously cluster all the granules iteratively. Each iteration will use results of previous clustering of connected granules, until a stable clustering of all the granules is achieved. In a hierarchical environment such as customers and products, it will mean that clustering of customers uses profiles of product clusters, and vice versa. For networked granules, a phone user is clustered using cluster profiles of the other connected users. In a temporal granular clustering, daily patterns will be clustered based on clustered profiles of historical and future patterns. These repeated applications of clustering are termed iterative in a hierarchy and are termed recursive in networks. The integrated meta-clustering of hierarchical, network, and temporal data is a multi-faceted project. Since clustering is unsupervised and we do not know the expected outcomes, it is important to study the quality of the resultant clustering. In addition to deriving quantitative evaluations, the notion of preference will be used to value a cluster based on how well-connected it is to more desirable objects. The iterative and recursive algorithms will be further modified for fuzzy and rough clustering, which allow an object to belong to multiple clusters. We plan to design, develop, implement, and test variations of the clustering algorithms for retail, mobile phone, engineering, and financial datasets.

SPEAKER

Pawan Lingras is a graduate of IIT Bombay with graduate studies from University of Regina. He is currently a professor at Saint Mary's University, Halifax and recently served as a UGC funded Scholar-in-Residence at SRTM University, Nanded and visiting professor at IIT Gandhinagar. He has authored more than 190 research papers in various international journals and conferences. He has also co-authored two textbooks, and co-edited two books and six volumes of research papers. His areas of interests include artificial intelligence, information retrieval, data mining, web intelligence, and intelligent transportation systems. He has served as the general co-chair, program co-chair, review committee chair, program committee member, and reviewer for various international conferences on artificial intelligence and data mining. He is also on editorial boards of a number of international journals.

**B. Title: Toward Granular Scalability of Analytical Data Processing**

**Speaker: Dominic Slezak**

ABSTRACT:

In this talk, we discuss Infobright's approach to interactive analytics over machine generated data. We focus on specific problems of carrying out investigative and predictive query workloads over rapidly changing datasets generated by systems, devices and sensors which are in turn the component pieces of large, emergent networks, such as those described as the Internet of Things. We present an architectural framework, referred to as Scalable Knowledge, for processing distributed data streams in order to compute standard or partially approximate query results. We explain how a network of so called knowledge processors can assist in analytical operations by propagating approximate summaries of granules of original and dynamically derived data. We also show how our approximate analytical model fits a more general idea of information granulation.

SPEAKER

Dominik Slezak received his D.Sc. (habilitation) in 2011 from Institute of Computer Science, Polish Academy of Sciences, and Ph.D. in Computer Science in 2002 from University of Warsaw, Poland. In 2005, he co-founded Infobright Inc., where he is currently working as chief scientist. He is also associate professor at Faculty of Mathematics, Informatics and Mechanics, University of Warsaw. He used to hold positions of assistant professor at University of Regina, SK, Canada, and in Polish-Japanese Institute of Information Technology in Warsaw.

Dominik serves as associate editor for several international scientific journals, including Information Sciences and Intelligent Information Systems. He is also in editorial board of Springer’s Communications in Computer and Information Science. He edited over 20 books and volumes of conference proceedings. He authored over 100 papers for books, journals and conferences. He delivered plenary talks at over 20 international conferences. His research interests include Rough Sets, Knowledge Discovery and Databases. In 2012-2014 he serves as president of IRSS.

**C. Title: Ocius, minutus, exactius – the Olympic Theme for Next Generation Machine Learning:**

**Speaker: Jayadeva**

ABSTRACT

Over the last decade and a half, support vector machines have become the paradigm of choice for most learning applications. However, new sources of data have emerged, ranging from high dimensional micro-array and bio-informatics data, to very large databases emanating from social networks and telecom service providers. The need to learn from such data presents new challenges, At the same time, the proliferation of handheld and portable computing and communication devices has motivated new applications that use recognition in novel ways, but which impose additional requirements on learning algorithms. This talk presents some of these challenges and will attempt to give a glimpse of some new directions.

Speaker

Jayadeva is a professor in the department of Electrical Engg., at IIT Delhi. He did his B.Tech and Ph.D from IIT Delhi. He works in the areas of Machine Learning, Neural Networks and VLSI design. He has received INSA Young scientist medal and Young Engineer award of INAE. His research work twin-SVM is widely cited and used.

**D. Title: Soft Computing Challenges in Algorithmic Trading**

**Speaker: Saugato Mukerji**

ABSTRACT

This presentation will bring forth “What Program Trading is?” It will deal with the major challenges that one encounters while developing trading models based only on stock-price data. We shall describe the major milestones that need to be traversed in the development of a working Trading model. Finally, it will also deal with reasons for which we have to take recourse to soft computing techniques for developing these algorithmic trading models and the present state of such development.

SPEAKER

Saugato is currently Vice President, Marketopper Securities Pvt. Ltd. He is heading the Research and Development team at Marketopper Securities Pvt. Ltd. He brings with him 12 years’ experience of development of different algorithm based trading Products as well as Risk Management Systems. He has a Bachelor’s Degree in Chemical Engineering as well as a Master’s Degree in Business Administration from A.M.U. Aligarh.

**E. Title: Persona Based Mobile Marketing**

**Speaker: Prateek Kapadia**

ABSTRACT

In this talk we shall examine use of pattern recognition techniques and soft computing for identifying abstract user profiles from large data base of millions of mobile users. This problem requires effective application of big data technology for machine learning problems. We shall discuss how practical services have been rolled out through this approach.

Speaker

Prateek is CTO of Flytxt dealing with technology challenges in mobile marketing in emerging and mature markets of Asia and Africa. He is also pursuing Ph.d from IIT Bombay.

**F. Title: Mining large document repositories**

**Speaker: Girish Keshav Palshikar**

ABSTRACT

Large repositories of text documents are now common, both on the Web as well as inside enterprise intra-nets. Efficiently extracting novel and actionable knowledge hidden in these

large document repositories is a prominent challenge, particularly from the Big Data perspective. Many machine learning techniques - such as clustering and classification - have been adapted for mining large text repositories. In addition, since unstructured text is quite different in nature from structured data, many peculiar problems crop up in mining text repositories. In this Talk, we will give a short overview of three problems in text mining and review some approaches to solve them: keyword extraction, named entity recognition and hyperlinking. We will also discuss some specific application scenarios such as mining large resume repositories.

SPEAKER

Girish Keshav Palshikar is working in the Tata Research Development and Design Centre (TRDDC), Pune, India, since 1992, where he is now a Principal Scientist and leads the Machine Learning R&D Group. He was honoured with the title of TCS Distinguished Scientist. TRDDC is one of the R&D and Innovation Laboratory of Tata Consultancy Services Limited, a premier software company in India. He has about 80 publications in international journals and conferences. He is also a visiting faculty at the Computer Science Department of University of Pune. His areas of research include machine learning, as well as data and text mining.

**G. Title: Cognitive Computing**

**Speaker: Raghavendra Singh**

ABSTRACT

Cognitive computing aims to develop a coherent, unified, universal mechanism inspired by the mind’s capabilities. Rather than assemble a collection of piecemeal solutions, whereby different cognitive processes are each constructed via independent solutions, we seek to implement a unified computational theory of the mind. IBM scientists are working on a project called SyNAPSE (Systems of Neuromorphic Adaptive Plastic Scalable Electronics), funded by the U.S. Defense Advanced Research Projects Agency (DARPA), which aims to develop electronic neuromorphic machine technology that scales to biological level and is inspired by the structure and architecture of the organic brain—the way neurons receive sensory input, connect to each other through synapses, adapt these connections, and transmit mental and motor output. This talk will introduce IBM's vision of Cognitive Computing, describe the design principles behind SyNapse project and discuss the applications that are being developed.

SPEAKER

Raghavendra Singh is a Research Staff member at IBM Research. He is currently a member of the Cognitive Computing group. His research interests are in in the area of information theory, signal processing and representation as applied to a spectrum of problems in neuro-science, compression and transmission of multimedia data, and monitoring of large scale data centers. He did his PhD. in Electrical Engineering from University of Southern California (2001). His undergraduate degree is from BITS Pilani, India (1993). He is a Senior Member of the IEEE.