IEEE CIS Summer School 2019

on

"Big Data Analytics and Stream Processing: Tools, Techniques and Application"

Location: Indian Institute of Information Technology Allahabad, India

When: August 10 - 14, 2019

Website: https://bdasp.iiita.ac.in/

Summary

In today's world, Big data analytics and stream processing have taken great hype due to the digitization of the environment along with the integration of smart data computing services and interconnectivity. This digitized world offers huge applications especially in the fields of agriculture, healthcare, smart education, economy, energy, industry, and a lot more. Most of the required data is gathered by the countless number of sensor devices being applied in the vicinity of humans to identify certain activities or scenarios. In order to process such a huge never-ending stream of data, there is a need to re-think the way data is processed for both cases i.e. data-at-rest and data-in-motion.

In order to have knowledge about the latest trend in this field, in this IEEE CIS Summer School on Big data analytics and stream processing is aimed at highlighting the tools, techniques, and applications from the perspective of future intensive applications. To serve this purpose, this IEEE CIS Summer School features a large number of keynote speakers/plenary/invited talks on advanced topics and also offers a good platform for the participants for the innovative and entrepreneurial ideas.

Committee

Patron

Prof. P. Nagabhushan, Indian Institute of Information Technology, Allahabad, India
https://www.iiita.ac.in/administration/director_profile/

Chair

Prof. Shekhar Verma (Honorary Chair), Indian Institute of Information Technology Allahabad,
https://it.iiita.ac.in/?pg=facultypage&uid=sverma

Dr. Sonali Agarwal (General Chair), Indian Institute of Information Technology Allahabad,
http://www.sonaliagarwal.com/

Co-Chair

Prof. Shirshu Varma, Indian Institute of Information Technology, Allahabad, India
https://it.iiita.ac.in/?pg=facultypage&uid=shirshu

Dr. Vrijendra Singh, Indian Institute of Information Technology, Allahabad, India
https://it.iiita.ac.in/?pg=facultypage&uid=vrij

Local Organising Committee

Dr. Pavan Chakraborthy, Indian Institute of Information Technology Allahabad, India
Dr. K. P. Singh, Indian Institute of Information Technology Allahabad, India
Dr. S. Venkatesan, Indian Institute of Information Technology Allahabad, India
Dr. Manish Kumar, Indian Institute of Information Technology Allahabad, India
Dr. Satish Kumar Singh, Indian Institute of Information Technology Allahabad, India
Dr. T Pant, Indian Institute of Information Technology Allahabad, India
Dr. Mohammed Javed, Indian Institute of Information Technology Allahabad, India

Acknowledgements

Our sincere thanks to:

- ✓ Prof. P. Nagabhushan for his support as BDASP'19 Tutorial Patron
- ✓ Prof. Shekhar Verma, Honorary Chair of BDASP'19 and his team
- ✓ IEEE UP SECTION INDIA and the IEEE CIS for making this summer school possible.

Schedule

Indian Institute of Information Technology Allahabad, Prayagraj

IEEE CIS Summer School 2019
Big Data Analytics and Stream Processing: Tools, Techniques and Application

August 10 - 14, 2019

Program Schedule

Time/Date	10.08.2019	11.08.2019	12.08.2019	13.08.2019	14.08.2019
8:30-9:00	Registration				
9:00-10:30	Inaugural Ceremony followed by tea	Prof. P. Nagabhushan (IIIT Allahabad, India)	Dr. Partha Pratim Roy (IIT Roorkee, India)	Prof. Kenji Doya (OIST, Japan)	Dr. Manish Kumar (IIIT Allahabad, India)
10:30–11:00	Dr. Shekhar Verma (IIIT Allahabad, India)				
11:00-11:30		Morning Tea Break			
11:30-12:30	Dr. Nischal K Verma (IIT Kanpur, India)	Dr. Sonali Agarwal (IIIT Allahabad, India)	Dr. R Krishnan (IIST, Kerala)	Prof. Vasudha Bhatnagar (DU, India)	Dr. Vrijendra Singh (IIIT Allahabad, India)
12:30-14:30	Lunch Break				
14:30-15:30	Prof. Ramesh K Agarwal (JLNU, India)	Dr. Rahul Kala (IIIT Allahabad, India)	Dr. Rajiv Mishra (IIT Patna, India)	Prof. O.P Vyas (IIIT Allahabad, India)	Rump Session
15:30-16:00	Evening Tea Break				
16:00-17:00	Prof. Hari Mohan (EHU, UK)	Prof. P. N. Suganthan (NTU, Singapore)	Prof. Kenji Doya (OIST, Japan)	Prof. O.P Vyas (IIIT Allahabad, India)	Closing Ceremony, City Tour and Departure
17:00-18:00	Dr. M. Tanveer (IIT Indore, India)				
18:30-20:30			Banquet Dinner		

Speakers

Incremental learning in the framework of Symbolic-Histogram objects for Big-Data

Prof. P. Nagabhushan

Conventional one shot learning, which is a one-go process with the entire dataset, fails to withstand the boom in data. Huge volumes of data have been compelling the learning methodology to shift towards processing smaller chunks of data followed by assimilating the knowledge in an incremental mode, which is identified as Incremental Learning. The phrase Incremental Learning refers to the process of deriving/updating knowledge in a phased manner without re-indenting the past data or with minimal re-indenting in unavoidable cases.

Incremental Learning provides a computationally better framework for i) Very large volume of data (ii) Temporally arriving data and (iii) Multi-source data. Since learning is by and large unsupervised, we have employed clustering as a tool for visualizing the incremental learning. Specifically, we have used Density Based Spatial Clustering of Applications of Noise (DBSCAN) algorithm, since the method remains insensitive to the order of presentation of data samples.

Two typical application case studies involving spatial data and multi-source temporal discrete data are considered to illustrate the customized utilization of the proposed SCIL and SOIL models for accomplishing the Incremental update of knowledge. An interesting application of tracking the outliers of each stage is also presented. As the basis for PCA is covariance matrix, the basis for Incremental PCA is Incremental covariance matrix. Formulation to incrementally update covariance matrix, particularly with temporal chunks of data expressed in terms of histogram and regression line features is successfully attempted. This has motivated us to take up deeper research in the direction of incremental dimensionality reduction and overall the research has given rise interest in us to take up further research in incremental learning.

Deep Learning with Fuzzy Systems

Dr. Nishchal K Verma

DFM has the capability to handle more complexity and abstraction in data with smaller architecture compared with DNN. Working of trained model is understandable to human beings and underlying working can easily be comprehended by human beings by just looking at network parameters.

DFN has the asset of dealing with uncertainty of various kinds such as vagueness, ambiguity, imprecision, etc. DFM is robust in presence of noise in data. Prior human supervisor knowledge can be easily incorporated into the architecture of DFN. It Closely matches with human cognitive thinking

Composite Kernel SVM in Conjunction with Spatial Filter for Brain Computer Interface

Prof. Ramesh K Agarwal

BCI is a system which takes a bio signal measured from a person and predicts some abstract aspect of the person's cognitive state. There exist diseases of the nervous system that gradually cause the body's motor neurons to degenerate Amyotropic Lateral Sclerosis (ALS). It eventually causes total paralysis. The affected individual becomes trapped in his own body, unable to communicate.

Utility of Genetic Algorithm and Grammatical Inference for Big Data Analytics: Challenges, Solutions and Applications in Big Data Analytics

Dr. Hari Mohan Pandey

Genetic Algorithms (GAs) are metaheuristic algorithm operates on encoding mechanism. GA's success depends on a good balance between exploration and exploitation. Exploration aims to visit entirely The key challenges with the metaheuristic algorithms including GAs are:

The aim of this talk is to provide a much deeper understanding about exploration and exploitation strategies and identify possibilities by which performance of the GAs can be improved. The domain of enquiry in this talk is grammatical inference (GI). GI is a methodology to infer context free grammars (CFGs) from training data and, it provides greater benefits in data mining and big data analytics. Through this talk an attempt is made to present a fresh treatment and discuss the utility of GA and GI for big data analytics in 4-rational aspects: (a) what are challenges in implementing a GA; (b) how to control exploration and exploitation in GA; (c) How to apply GA for real word problems and, (d) How GA and GI are suitable for big data analytics. The obvious outcome of this talk is to create an awareness among the researchers about the suitability of GA and GI for big data analytics. I also believe that this talk might be useful for the researcher to identify the possibilities to develop new algorithm in the big data analytics domain.

Large scale least squares support vector machines

Dr. M. Tanveer

Twin support vector machine (TSVM), least squares TSVM (LSTSVM) and energy-based LSTSVM (ELS-TSVM) satisfy only empirical risk minimization principle. Moreover, the matrices in their formulations are always positive semi-definite. To overcome these problems, we propose in this paper a robust energy-based least squares twin support vector machine algorithm, called RELS-TSVM for short.

Unlike TSVM, LSTSVM and ELS-TSVM, our RELS-TSVM maximizes the margin with a positive definite matrix formulation and implements the structural risk minimization principle which embodies the marrow of statistical learning theory.Furthermore, RELS-TSVM utilizes energy parameters to reduce the effect of noise and outliers. Experimental results on several synthetic and real-world benchmark datasets show that RELS-TSVM not only yields better classification performance but also has a lower training time compared to ELS-TSVM, LSPTSVM, LSTSVM, TBSVM and TSVM.

Data Stream Analytics for Cyber Physical Systems

Dr. Sonali Agarwal

Streaming data are potentially infinite sequence of incoming data at very high speed and may evolve over the time. This causes several challenges in mining applications of Cyber Physical Systems (CPS) due to processing needs of large scale high speed data streams in real time. Hence, this field has gained a lot of attention of researchers in previous years.

The present research work discusses various challenges associated with mining such data streams especially in the domain of Cyber Physical Systems. Several available stream data mining algorithms of classification and clustering are specified here along with their key features and significance. Also, the significant performance evaluation measures relevant in streaming data classification and clustering are important to observe with their comparative analysis. The research work also illustrates various streaming data computation platforms that are developed and their significance in cyber physical systems and discusses each of them chronologically along with their major capabilities. It also clearly specifies the potential research directions open in high speed large scale data stream mining from algorithmic, evolving nature and performance evaluation measurement point of view.

Making sense of High Rate/High Definition Data in Self-driving Cars

Dr. Rahul Kala

The self-driving cars are equipped with a large number of high definition sensors like 3-D lidars and stereo cameras. The primary necessity is to make real time decisions. In this talk we look at the most important module to calculate accurate the position of the vehicle based on the sensor readings, while also representing the rich sensor information as a map, commonly known as the problem of Simultaneous Localization and Mapping.

The talk tours around the beautiful campus and illustrates the technology that can automatically make a map of the campus so that next time the vehicle can know where it is, just by a sight using the sensors. The talk also looks into the possibilities to learn the

localization and navigation modules using machine learning on such a high definition data. Lastly, the talk delves into making the decisions under dynamic and adversarial settings.

Application of Machine Learning using Sensors

Dr. Partha Pratim Roy

Over the last years, information technology (IT) has moved very rapidly from desktop to mobile computing. With the advanced and extensive use of smartphones, smart watches and head-mounted devices, this paradigm shift in terms of computing from home-office environment to an anytime-anywhere activity.

Introduction of low-cost depth sensors such as Leap motion or Kinect has made capturing of 3D data more convenient. These sensors are used to capture finger and gesture information quite effectively. In the first part of my talk, the use of machine learning for developing applications will be demonstrated. It will include applications like, document image processing, 3D-air Gestural User Interfaces, 3D air writing, etc. In the second part of my talk, I will present applications of EEG technology on brain-computer-interface (BCI). The development of Electroencephalography (EEG) sensor technology through wireless headsets and their connectivity with mobile devices has opened-up new ways of implementation of Brain Computer Interface (BCI) applications. Traditionally, EEG signals are used in diagnosing diseases. Since brain signals represent the physiological and behavioural information about a person, these signals are widely used in developing biometric, predictive, emotion and gaming applications. In this talk, analysis of EEG signals and few applications of EEG technology beyond clinical applications using brain-computer-interface (BCI), will be presented.

Security for Images and Images for Security

Dr. R Krishnan

Images can be transmitted with embedded covert information. This can be for legitimate or malicious purposes. This is referred to as Steganography. Alternately Image fragments can be used to send/store information which can then be retrieved securely.

This process is called Visual Cryptography. The first half of the title of this talk refers to Steganography and the second to Visual Cryptography. Steganography is primarily used for the covert transmission of information even though the purpose can be legitimate or malicious. It has been used from historical Times. In this talk I will discuss the main methods of steganography like LSB embedding etc. The methods of steg analysis which is the reverse process of examining whether steg content is present will also be briefly explained. The method to build a steg firewall which will filter steg content without identifying it explicitly forms the basis of a thesis by one of my students. This will be described Along with the results. Briefly the method involves the use of low level radiometric and geometric image processing operations on the steg image. This has the effect of smearing the steg content and will largely make them unintelligible. At the same time the cover image will also get degraded. A deconvolution process is proposed which will restore the cover image quality largely.

The PSNR of the restored images was in the acceptable range of 30-40. Using steg analysis methods the obfuscation of the steg content is also confirmed. In visual cryptographic schemes shares of secret images look like random patterns when stacked together they produce meaningful images. By distributing these shares or by sending them through different channels security can be ensured. The basic VCS scheme will be described and the extension proposed by one of my students will also be discussed. In the EVCS scheme the shares look like meaningful images. In the case of ideal contrast deterministic construction for VCS, each participant needs to hold one or multiple image shares with same size of the binary secret image and the secret image will be reconstructed without any change in resolution.

Storage technologies for Big Data

Dr. Rajiv Mishra

The new generation of storage systems called key value stores as the storage systems for Big Data. We will see the storage technologies for big data concept and the architecture of different systems. These big-data storage technologies are used by companies like Twitter comm where you receive millions of tweets from millions of users you might maintain a key value store, the online retailer like Amazon also maintain the information about item etc, the kayak.com flight booking online store where the flight information.

The big data involved in such online apps need to maintain this data on a distributed cluster of servers i.e. the database maintains large amounts of data and this data can then be queried it can be looked up and so the companies like Netflix. We will discuss at the design of a real key value store Apache Cassandra which is one of the most popular key value stores that is being used in industry today.

Artificial Intelligence and Brain Science and Big Data Challenges in Neuroscience

Prof. Kenji Doya

Human brain functions have long served as the targets of development of artificial intelligent systems. Findings in neuroscience have provided guidance at multiple levels in the designs of machine learning architectures and algorithms. Today's neuroscience, in turn, necessitates applications of artificial intelligence and machine learning algorithms for making sense of huge datasets. This lecture reviews examples of co-evolution of AI and brain science and consider how they can help each other for further progress.

Today's neuroscience is becoming data science, thanks to progresses in high-throughput instrumentation, including two-photon calcium imaging, serial section electron microscopes, and single-neuron RNA sequencing. This lecture first reviews computational methods used for big brain data analysis and modelling. Examples are then presented in calcium imaging data analyses for identifying the neural circuits for reinforcement learning and mental simulation.

F.A.T.E. in Artificial Intelligence

Prof. Vasudha Bhatnagar

Big data, augmented with theoretical breakthroughs in Machine Learning and hardware advances, has been the primary driver of Artificial Intelligence research, development and deployment. Touched by AI technology on personal devices, humanity is heading towards an environment where critical and crucial decisions will be taken by Artificial Intelligence.

My talk will touch the concerns that are being raised over the deep penetration of artificial intelligence in Healthcare, Finance, Education, Recruitment, Autonomous driving vehicles, Robotics etc. The dual goal of this talk is to sensitise the school attendees towards development of responsible AI, and the educators to formally include content regarding these concerns in Ai courses.

From Predictive to Prescriptive Analytics: Challenges & Opportunities

Prof. O.P. Vyas

Though Data Analytics has become integral part of the decision making in the most enterprises but still approaches are mostly either descriptive or predictive in nature. Whereas the decision points / action points are not precisely formed in these approaches and also the associated costs-benefits are not thoroughly discussed in the current scenario.

Prescriptive approaches are being explored to fill the gaps and is often considered as the next step towards increasing data analytics maturity and leading to optimized decision making ahead of time for business performance improvement.The prescriptive Analytics is though relatively new and approaches and performance metrics thus poses interesting challenges to be addressed by the researchers. The proposed talk while presenting the state of art in the prescriptive analytic approaches, addresses the related issues with specifics associated with the Prescriptive Analytics framework for an improved decision making.

Big/Large data mining-Algorithms

Dr. Manish Kumar

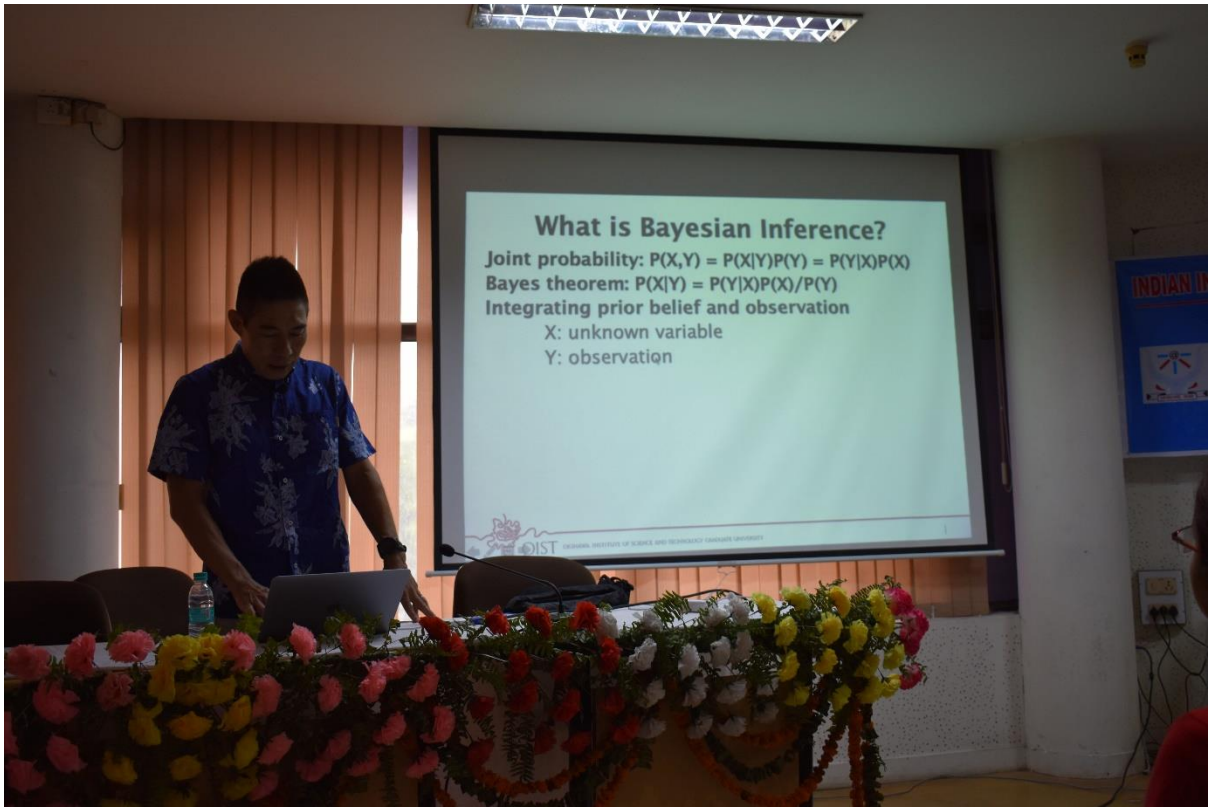
This talk will focus on distributed and parallel algorithms of data mining. Important methods and algorithms for big and large mining will be covered. Research issues and challenges will be discussed. Association mining and clustering algorithms using Map Reduce implementation will be covered.

Time series data analysis

Dr. Vrijendra Singh

Time series analysis has been always important for pattern recognition researchers. Time series data analysis comprises methods for analysing time series data in order to extract meaningful information for the purpose of modelling, classification, clustering and forecasting etc. Parametric and non-parametric techniques have been widely used in the study of time series data. Recent advancements like time series anchored chains and deep learning model have been presented including real life applications and challenges.

Photos


What is Bayesian Inference?
Joint probability: P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X)
Bayes theorem: P(X|Y) = P(Y|X)P(X)/P(Y)
Integrating prior belief and observation
X: unknown variable
Y: observation


REGISTRATION


भारतीय सूचना प्रौद्योगिकी संस्थान, इलाहाबाद


14.08.2019 12:49


Computational
Intelligence
Society
Uttar Pradesh Section
14.08.2019 12:49