GLOBAL INITIATIVE OF ACADEMIC NETWORKS (GIAN)

Ministry of Human Resource Development
Government of India

Parallel and Distributed Data Stream Mining

Overview

Data is being continuously collected from a variety of sensor sources, such as Twitter feeds, news streams, and environmental sensors. It is a significant challenge to continuously monitor such data and derive insights in a timely manner. This course on data stream analysis focuses on methods and software for deriving patterns and aggregates from data streams in real-time. The course will focus on (1) Parallel and distributed methods for data stream mining, (2) Methods for mining from graphical data, where each stream item represents a relationship between entities.

Objectives

The main objectives of the course are:

1. Introduce the student to use cases of stream processing, the data stream model and graph stream model
2. Present algorithmic techniques for graph stream processing, including random sampling, graph sketches, and merge-and-reduce.
3. Show their application to problems such as subgraph counting, graph connectivity, random sampling from graphs, graph matchings, etc
4. Present current techniques on monitoring parallel and distributed streams, including algorithms in the continuous distributed monitoring model, and the parallel streaming model
5. Provide practical perspective on building software for stream processing
6. Provide experience with an open source tool, Apache Flink

Teaching Faculty with allotment of Lectures and Tutorials

1. Prof. Srikanta Tirthapura (ST) : 8 hrs lectures and 5 hrs tutorials
2. Prof. Sonali Agarwal (SA): 5 hrs lectures and 4 hrs tutorials

Duration: Dec 19 – Dec 23, 2017: 13 hrs lectures and 10 hrs Tutorials

Course Details

Day 1: Basic Fundamentals of Stream Processing

Lecture 1: 1:30 hrs. (10:00 am to 11:30 am): ST
Fundamentals of Stream Processing, Models, Algorithms, and Systems

Lecture 2: 1:30 hrs. (12:00 pm to 1:30 pm): SA
High Velocity Data Stream mining algorithms & techniques: I

Tutorial 1: 2 hrs. (3:30 pm to 5:30 pm): ST
Basics of Stream Processing Software using Apache Flink

Day 2: Data Stream Sampling and Mining

Lecture 3: 1:30 hrs. (10:00 am to 11:30 am): SA

High Velocity Data Stream mining algorithms and techniques: II

Lecture 4: 1 hrs. (12:00 pm to 1:00 pm): ST

Random Sampling from Data Streams

Tutorial 2: 2 hrs. (3:30 pm to 5:30): ST

Processing Streaming Data Using an Operator Pipeline

Day 3: Graph Stream Processing

Lecture 5: 1:30 hrs. (10:00 am to 11:30 am): ST

Graph Stream Processing: I

Lecture 6: 1 hrs. (12:00 pm to 1:00 pm): ST

Graph Stream Processing: II

Tutorial 3: 2 hrs. (3:30 pm to 5:30):SA

Distributed stream processing using Apache Flink

Day 4: Distributed/Parallel Stream Processing and Challenges

Lecture 7: 1:30 hrs. (10:00 am to 11:30 pm):ST

Parallel and Distributed Stream Processing

Lecture 8: 1 hrs. (12:00 pm to 1:00 pm): SA

Performance Measures and Challenges of Data Stream

Tutorial 4: 2 hrs. (3:30 pm to 5:30):SA

Use cases of real-time Data Stream Mining

Day 5: Stream Learning and Complex Event Processing

Lecture 9: 1:30 hrs. (10:00 am to 11:30 am): ST

Machine Learning on Data Streams

Lecture 10: 1 hrs. (12:00 pm to 1:00 pm):SA

Complex Event Processing on Data Streams

Tutorial 5: 2 hrs. (3:30 pm to 4:30):ST

Rump session for participants

Who can attend

- Faculty, engineers, scientists, and researchers from academic, industrial and government organizations including R&D laboratories from India or abroad.
- Students at all levels (BE/BTech/MSc/ME/MTech/PhD/Other) from academic and technical institutions/universities from India or abroad
- Number of participants for the course will be limited to fifty. Preference will be given to the participants opting against credits.

Fees

- The participation fees for taking the course is as follows:
  - Participants from abroad : US $500
  - Industry/Research Organizations: INR 10,000
  - Academic Institutions: INR 2,000 (half for SC/ST students)

The above fee includes all instructional materials, tutorials and assignments, 24 hours free internet facility. The participants will be provided with campus accommodation on payment basis.

Contact person

Dr. Sonali Agarwal
Assistant Professor,
Department of Information Technology,
Indian Institute of Information Technology, Allahabad
Devghat, Jhalwa, Allahabad 211015
E-mail: sonali@iiita.ac.in
Mobile:+91-9415647042

Course Registration Process

Step 1: One Time Registration

Registration for GIAN courses is not free because of constraint in the maximum number of participants allowed to register for a course.

In order to register for any course under GIAN, candidate will have to get registered one time first to GIAN Portal of IIT Kharagpur using the following steps:

I. Create login and password at http://www.gian.iitkgp.ac.in/GREGN/index

II. Login and complete the Registration Form

III. Select Courses

IV. Confirm your application and payment information

V. Pay Rs. 500/- (non-refundable) through online payment gateway

VI. Download and print "pdf file" of your enrolment application form for your personal records and copy of the same to be sent to the Course Coordinator.

Step 2: Institute Registration

I. Institute registration process is an online process through Google Forms. Shortlisted candidates will be provided to fill this form for submission to the Coordinator at appropriate time.

Course Fee (Non-refundable): The participation fee to attend the short course shall be:

Participants from abroad:	US $500
Industry/Research Organizations:	INR 10,000
Academic Institutions:	INR 2,000 (half for SC/ST students)

The above fee includes the instructional materials, internet facility and snacks between the sessions. The accommodation will be provided on payment basis subject to availability on request otherwise participants will have to make their own stay arrangement.

II. The Registration fee has to be paid through National Electronic Funds Transfer (NEFT) to the account of "Indian Institute of Information Technology, Jhalwa Allahabad" (Account No. : SB A/c. 35465510718, SBI, Branch Jhalwa, Allahabad, Branch Code: 10891, MICR Code: 211002057, IFSC Code: SBIN0010891.

III. Scan copy of the Bank Transfer Certificate "Receipt" must be sent via Email to the GIAN

Coordinator of the gian@iiita.ac.in on or before November 30, 2017.

IV. Postal address for any correspondence, though email is preferable.

Dr. Sonali Agarwal
Assistant Professor,
Department of Information Technology,
Indian Institute of Information Technology, Allahabad
Devghat, Jhalwa, Allahabad 211015
E-mail: sonali@iiita.ac.in
Mobile:+91-9415647042

V. Selection will be made purely on First Come First Serve basis according to eligibility and subject to fulfilling of the available seats.

VI. Maximum fifty (50) participants will be accommodated in the course. 8. The Brochure and the Registration Form may be downloaded from the Institute website https://gian.iiita.ac.in.

Important Information

I. The students will obtain academic credits for this course, if opted, based on the evaluation rules and grading process of the GIAN and IIITA.

II. After successful completion of the course, all participants will get participation certificates along with grades and credits, if registered against credits, as per Institute norms.

III. No TA, DA will be provided to the participants.

IV. Limited accommodation is available in the Institute campus which will be provided on First Come and First Serve Basis on payment mode.