Global Initiative of Academic Networks(GIAN)

course on

Data Mining Algorithms for Graph, Text, and Streaming Data

December 19-27, 2017

at

Computer Science and Engineering Discipline

PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, MP, India

---

About GIAN

A new program titled "Global Initiative of Academic Networks (GIAN)" was initiated by Govt. of India in Higher Education aimed at tapping the talent pool of scientists and entrepreneurs, internationally to encourage their engagement with the institutes of Higher Education in India so as to augment the country's existing academic resources, accelerate the pace of quality reform, and elevate India's scientific and technological capacity to global excellence.

Under the GIAN program lectures by internationally and nationally renowned experts are being arranged to garner the best international experience into our system of education, enable interaction of students and faculty with the best academic and industry experts from all over the world and share their experiences and expertise to motivate people to work on Indian problems.

About PDPM IIITDM Jabalpur

IIITDM Jabalpur was established in 2005 with a focus on education and research in IT enabled Design and Manufacturing. Since its inception, PDPM IIITDM Jabalpur has been playing a vital role in producing quality human resources for contribution in India's mission of inclusive and sustainable growth. The Institute offers undergraduate, post graduate and PhD programmes in Computer Science and Engineering, Electronics and Communication Engineering, Mechanical Engineering, Design, and PhD programmes in Mathematics and Physics. Under IIIT act, the Institute has been declared an Institute of National Importance in January 2015.

Course Overview:

Analysis of large amounts of data is becoming crucial for most spheres of our economy and society. Scientific research, engineering practices, and business operations are all becoming increasingly dependent on automated analysis of large repositories of digital data. The digital data in today's contexts is becoming very large in size and is much richer in its representation, structure, and semantics, when compared to a collection of numeric values. Most of the data being collected today is in the form of collections of text documents, images, networks and graphs, and time series. There are complex spatial, temporal, and/or other structural interrelationships among data elements, including flow parameters along edges of a graph and grammar structures underlying text documents.

The proposed course will provide some insights into data analysis challenges for complex data types of graphs, texts, and data streams. The course will provide an outline of recent developments in the applications for these data types and research results. The traditional algorithms for classification and clustering have seen novel adaptations and also completely redesigned to meet the needs of these complex data types. The traditional problems of clustering data points have evolved to address similar problem in graphs, texts, and data streams with the goal of discovering interesting communities, pathways and nodes in social and/or scientific networks. Statistical methods have been adapted to handle streaming datasets in which we can only store data from small intervals and statistical assumptions of large sample sizes cannot be satisfied. Recommendation and anomaly detection systems have become very important due to their valuable applications in business and scientific domains. The proposed course will provide some deep insights into the nature of analysis problems faced by these complex types of datasets.

The course Instructor:

Prof. Raj Bhatnagar is Professor of Computer Science at University of Cincinnati, Ohio, USA. His area of research is data mining and pattern recognition. He has worked on problems in this research area for more than twenty five years. His research projects have been funded by NSF, US Air Force, US DARPA, and a number of Industrial sponsors. His recent research projects include design of mining and analysis algorithms for Big Data situations in Biomedical, Manufacturing, GIS, and Security applications. These problems have involved various types of structured and unstructured data. He has published more than eighty peer-reviewed publications. He has designed and taught graduate level classes on the topics of Data Mining, Big Data Analysis, and Artificial Intelligence.

Raj K Bhatnagar, Professor of Computer Science
Department of Electrical Engineering and Computing Systems
University of Cincinnati, Cincinnati, OH 45221, USA
Raj.Bhatnagar@uc.edu +1 (513) 556-4932

Global Initiative of Academic Networks

a course on

Data Mining Algorithms for Graph, Text, and Streaming Data

December 19-27, 2017

at

Computer Science and Engineering Discipline

PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, MP, India

---

Who can attend:

1. Students pursuing BTech/MTech/MS/MSc/Ph.D. degrees in computer and information sciences or Faculty from reputed academic and technical institutions.
2. Executives and researchers from government/non-government organization working in various domains of Sciences, Engineering, and Business Analytics who need to process complex and large datasets.

You should attend if:

- You wish to get familiarity with a number of problems from the sciences, engineering, and business domains that generate large and complex data and are in need of knowledge discovery algorithms.
- You wish to get theoretical understanding and hands-on experience with some existing algorithms for clustering, classification, and association analysis of simpler data types.
- You wish to see adaptation of mining algorithms for the more complex and structure rich data types, specifically, graph clustering, formal concept analysis, text mining, and stream data processing.
- You wish to get an understanding of complexity and scalability aspects, especially relating to the processing of extremely large data sets is parallel environments using Hadoop and Map-Reduce types programming environments.

Important Dates:

Last Date of Online Registration: December 10, 2017

Course Dates: December 19-29, 2017

Registration Steps

1. Register online at: http://www.gian.iitkgp.ac.in/GREGN/index
2. The registration fee can be paid through a Demand Draft drawn in favour of "PDPM IIITDM JABALPUR" payable at JABALPUR OR Through NEFT/RTGS :

Account Name:	PROJECT ACCOUNT PDPM IIITDM JABALPUR
Account No.:	50210022387
Bank MICR Code:	482010013
Bank IFS Code:	ALLA0212433
Bank Name:	Allahabad Bank
Branch Name:	Mehgawan, IIITDM, Campus Branch, Jabalpur

3. Please add the DD number or transaction number in the Google following form.
   https://docs.google.com/forms/d/e/1FAIpQLScc0SV3bPN76MUCWHX-ydTLLJa1OwOjg2JPi1SgOUE-B-HmxA/viewform
4. Please email transaction details and registration copy (from GIAN) to Dr. Pritee Khanna, Email: datamining@iiitdmj.ac.in

Registration Fee

Industry/ Research Organizations: INR 10000
Academic Institutions (Faculty): INR 6000
Research scholars: INR 4000
Students: INR 3000
Participants from abroad: US $250
The registration fee includes instructional materials, tutorials and assignments, computer and Internet access, food and accommodation in institute hostels. However, if accommodation is not required then INR 1000 would be refunded back.

Global Initiative of Academic Networks

a course on

Data Mining Algorithms for Graph, Text, and Streaming Data

December 19-27, 2017

at

Computer Science and Engineering Discipline

PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, MP, India

Schedule:

Lectures/ Tutorials	Day	Date	Topics
Lecture-1	2 Hours	Dec 19, 2017	Data Types and their characteristic features and properties. Graphs, high dimensional binary and real valued datasets; text collections and text feature types; streaming data and its problems from the perspective of data mining.
Tutorial-1	2 Hours	Dec 20, 2017	Write simple code to extract primitive features of each of the data types covered in Lecture-1.
Lecture-2	2 Hours		Main types of clustering algorithms, their basic properties and output characteristics. Hierarchical, sequential, partitional, and density based approaches
Lecture-3	2 Hours	Dec 21, 2017	Adaptation of clustering algorithms for graphs, and text data collections. Novel clustering algorithms for text analysis such as authorship attribution.
Tutorial-2	2 Hours		Programming exercises for clustering algorithms introduced in Lecture-2 and Lecture-3
Lecture-4	2 Hours	Dec 22, 2017	The notions of biclusters in large and sparse binary datasets and co-clustering for binary and real-valued datasets; algorithms for such clustering; correspondence between biclusters and cliques of graph structures. Formal concept analysis and its relationship to knowledge discovery.
Tutorial-3	2 Hours		Biomedical and other applications for biclustering. Biomedical datasets and text analysis datasets will be shown as example cases where these algorithms are useful.
Tutorial-4	2 Hours	Dec 23, 2017	Programming exercises for various types of clustering introduced in Lecture 4.
Lecture-5	2 Hours		Text mining problem and algorithms. Examples from text collections and types of results that can be obtained will be discussed. Authorship analysis, topic modeling, and clustering of text documents.
Tutorial-5	2 Hours	Dec 24, 2017	Programming exercises for text mining and analysis examples.
Lecture-6	2 Hours		Streaming data peculiarities and algorithms for mining streaming data.
Lecture-7	3 hours	Dec 26, 2017	Issues relating to processing of graphs in hadoop environments using Map-Reduce paradigm.
Lecture-8	3 hours	Dec 27, 2017	Applications and various research issues relating to large and complex data mining followed by open discussion.

Coordinators:

Dr. Pritee Khanna is currently working as an Associate Professor and head computer science & Engineering Discipline in Indian Institute of Information Technology, Design and Manufacturing Jabalpur, MP, India. She received her Ph.D. degree in computer science from Kurukshetra University, Kurukshetra, India in 2004. She is recipient of Long Term JSPS Fellowship and worked at Tokyo Institute of Technology, Japan. Her current research focuses on designing algorithms for image retrieval, biometrics, gesture recognition and medical image processing.

Tel(O): +91-761- 2794222

email: pkhanna@iiitdmj.ac.in and priteekh@gmail.com

Dr. Sraban Kumar Mohanty is currently working as an Assistant Professor in computer science & Engineering discipline in Indian Institute of Information Technology, Design and Manufacturing Jabalpur, MP, India. He received his Ph.D. degree in computer science and engineering from Indian Institute of Technology Guwahati, Assam, India in 2010. His current research focuses on designing of clustering algorithms for massive data sets using Eigen analysis.

Tel(O): +91-761-2794224

email: sraban@iiitdmj.ac.in and sraban@gmail.com