Unsupervised Data Mining: From Batch to Stream Mining Algorithms

....

Overview

Data mining is a field of research of computer science which is concerned with extracting patterns and models from data. Within this area of research a number of different topics are actively pursued. Supervised methods primarily focus on the prediction of discrete classes (classification) and numerical values (regression). Unsupervised methods, in contrast, derive statistical information from the data in the form of dependencies among the variables, groups of observations, distributions and processes that drive the generation of the data. In this course, we focus on methods for unsupervised data mining: (i) Pattern mining methods aim at discovering relevant patterns in data to explain or predict phenomena. (ii) Clustering is the task of grouping similar data in the same group (cluster). It aims to find clusters with high intra-cluster similarity and low inter-cluster similarity. (iii) Density estimation is the task of statistically estimating the density of data, i.e. of learning the joint probability distribution of variables of interest. Finally, (iv) process mining is the task of finding human-comprehensible descriptions of processes from data that are given as event logs.

Methods for unsupervised data mining were originally developed for processing batches of data. However, data analysis methods today have to cope with increasing amounts of data that often arrive in a stream. This means that the data are continuously produced. The same tasks that are performed on batches of data can also be performed on data streams including the above unsupervised data mining tasks: pattern mining, clustering, density estimation and process mining. The main difficulty in handling data streams lies in the speed of the arriving data and therefore the large amount of data that has to be processed. Saving all data and processing it later is infeasible when dealing with such large amounts. Therefore, techniques have to be developed that process the data one by one or in smaller batches, continuously updating patterns or models with new information. Other challenges are the possible occurrence of drift, the emergence of new trends and concepts, but also the repeated/cyclic occurrence of distributions. The objective of the course is to give a detailed description of different concepts related to batch and stream mining. It will provide a great opportunity for young researchers of data mining to learn more about state-of-the-art concepts of unsupervised stream mining.

Modules	**Unsupervised Data Mining: From Batch to Stream Mining Algorithms: September 24 - September 28, 2018** **Number of participants for the course will be limited to fifty.**
You Should Attend If…	▪ you are an executive, engineer and researcher from service and government organizations including R&D institutions. ▪ you are a student at any level (BTech/MSc/MTech/PhD) or Faculty from a reputed academic institution or technical institution interested in the state of the art in data mining.
Fees	The participation fees for taking the course are as follows: **Participants from abroad : US $500** **Industry/ Research Organizations: Rs. 10000/-** **Academic Institutions: Rs. 5000/-** **Students: Rs. 1000/-** The above fees include all instructional materials, computer use for tutorials, 24 hr free internet facility. The participants will be provided accommodation on payment basis.

The Faculty

Prof. Stefan Kramer is full professor, head of department of the institute of computer of Johannes Gutenberg University (JGU) Mainz, Germany, and honorary professor of The University of Waikato in Hamilton, New Zealand. Before his appointment at JGU, he was associate professor at the computer science department of Technische Universität München (2003 to 2011). He has been active in the field of data mining since the first conference worldwide in 1995 and is author of award-winning papers at ICDM, KDD, ILP and ICBK. He was vice-chair of ICDM 2013 and is regularly area chair of conferences like ECML/PKDD. His research interests include mining structured data, stream mining, process mining and clustering.

Prof. Pushpak Bhattacharyya is a Professor of Computer Science and Engineering at IIT Bombay. He is currently the Professor and Director of IIT Patna. He was a Visiting research fellow in Massachusetts Institute of Technology in 1990. He is well known for his contributions to natural language processing and has several distinctions in that field. Prof. Bhattacharyya specializes in Natural Language Processing (NLP), Machine Learning, Machine Translation, Cross Lingual IR and Information Extraction. A highly recognized researcher, Prof. Bhattacharyya is a member of National Knowledge Commission. He was the president of ACL (highest body of computational linguistics). For his outstanding contribution to computer science, he has received many accolades including IBM Innovation award (2007), Yahoo Faculty Award (2011), P. K. Patwardhan Award for Technology Development (2008), and VNMM Award of IIT Roorkee (2014).

Dr. Sriparna Saha is currently a Faculty Member of the Department of Computer Science and Engineering, Indian Institute of Technology Patna, India. Her current research interests include pattern recognition, multiobjective optimization and biomedical information extraction. Her h-index is 19 and total citation count of her papers is more than 2000 (according to Google scholar). She is the recipient of the Lt Rashi Roy Memorial Gold Medal from ISI Kolkata, Google India Women in Engineering Award, 2008, Junior Humboldt Research Fellowship, NASI Young Scientist Platinum Jubilee Award 2017, BIRD Award 2017, IEI Young Engineers' Award 2017 etc.

Dr. Asif Ekbal is currently a faculty member of the department of Computer Science and Engineering, Indian Institute of Technology Patna, India. His current research interests include Natural Language Processing, Data Mining and Machine Learning. His h-index is 23 and total citation count of his papers is more than 1800 (according to Google scholar). He is the recipient of the *Best Innovative Project Award* from the INAE, best paper award in ICACCI 2012, JSPS Invitation Fellowship from the Govt. Of Japan and Visvesvaraya Young Faculty Research Fellowship Award from Govt. of India.

Course co-ordinators

Professor Pushpak Bhattacharya
Director, Indian Institute of Technology Patna
Vijay and Sita Vashee Chair Professor
Department of Computer Science and Engineering
Indian Institute of Technology Bombay
Email: pb@iitp.ac.in
http://www.cse.iitb.ac.in/~pb

Dr. Sriparna Saha
Assistant Professor
Department of Computer Science and Engineering
Indian Institute of Technology Patna
Bihta, Kanpa Road, District – Patna
Bihar, India – 801118
Ph: +91-8809559190(m)
Email: sriparna@iitp.ac.in
http://www.iitp.ac.in/~sriparna/

Dr. Asif Ekbal
Associate Professor
Department of Computer Science and Engineering
Indian Institute of Technology Patna
Bihta, Kanpa Road, District – Patna
Bihar, India – 801118
Ph: +91-8809559190(m)
Email: asif@iitp.ac.in
http://www.iitp.ac.in/~asif/