High-Performance Data Mining: An Essential Paradigm for Knowledge Discovery

Prof. Alok N. Choudhary
John G. Searle Professor and Chair, Electrical Engineering and Computer Science
Northwestern University

Knowledge discovery in science and engineering has been driven by theory, experiments and more recently by large-scale simulations using high-performance computers. Modern experiments and simulations involving satellites, telescopes, high-throughput instruments, imaging devices, sensor networks, accelerators, and supercomputers yield massive amounts of data. At the same time, the world of business, traditional and online, as well as social communities are creating massive amounts of data at an astonishing pace. Thus high-performance data analytics and mining has become essential for knowledge discovery in science and engineering, as well as a business intelligence mechanism for operational success and sustainable competitive advantage in business.

In this talk, followed by an introduction to high-end data mining and the basic knowledge discovery paradigm, I will present research in high-end data mining along with live demonstrations. The approach in this talk would be application and example driven. We present (1) Data Mining techniques to mine massive web datasets from customer reviews for sentiment analysis and online recommendations; (2) Real-time stream mining of text from millions of blogs ad tweets to identify influencers and sentiments of people; (3) Predictive modeling using medical history data and data using devices for early breast cancer detection, survivability analysis and others; (4) Mining of traffic sensor data to predict patters and find associations; and (5) Other examples including mining in climate prediction, identification of protein interaction using text mining, mining of simulation data for identifying clusters in cosmological simulations.

Biography:

Alok Choudhary is a John G. Searle Professor and chair of Electrical Engineering and Computer Science at Northwestern University. He is the founding director of the Center for Ultra-scale Computing and Information Security (CUCIS). Prof. Choudhary was a co-founder and VP of Technology of Accelchip Inc., in 2000, which was eventually acquired by Xilinx. He received the National Science Foundation's Young Investigator Award in 1993. He has also received an IEEE Engineering Foundation award, an IBM Faculty Development award, an Intel Research Council award.
He is a fellow of IEEE, ACM and AAAS. His research interests are in high-performance computing, data intensive computing, scalable data mining, computer architecture, high-performance I/O systems and software and their applications in many domains including information processing (e.g., data mining, CRM, BI) and scientific computing (e.g., scientific discoveries). Alok Choudhary has published more than 350 papers in various journals and conferences and has graduated 27 PhD students. Techniques developed by his group can be found on every modern processor and scalable software developed by his group can be found on most supercomputers. Alok received his Ph.D. degree in Electrical and Computer Engineering from the University of Illinois, Urbana-Champaign, in 1989.