Subspace Clustering in Multi-Domain Datasets

Raj Bhatnagar
Computer Science Department
University of Cincinnati

Tuesday, April 14, 2009
3:00pm Rm. 110 Purdy-Kresge Library

Abstract:
Clustering provides significant insights into the structures embedded in a dataset. High Dimensional and sparse datasets have been dealt with using subspace clustering algorithms to determine interesting concepts in the data. The situation of multiple datasets, each relating two of the domains, presents a challenge that can provide much needed insights by scientists and business domain practitioners. For example, we may have a gene-diseases dataset, a gene-drug dataset, a drug-adverse reactions dataset, and a gene-pathways dataset. We would like to discover subsets from multiple domains, in the form of a multi-domain complex, that are related to each other and provide interesting insights into the datasets. We present our completed and ongoing research in the area of discovering such multi-domain clusters from datasets that contain binary data. We show the applicability of such algorithms and challenges that need to be addressed in generalizing these algorithms to the datasets containing real values.

Biography:
Raj Bhatnagar is an Associate Professor of Computer Science at University of Cincinnati. His main research interests are in data mining and clustering algorithms and his work has been supported by various Federal and State government agencies. He has applied the results of his research, over the years, to radar signatures, geographical data, and most recently to genomic data.