

Computer Science Seminar Series, 2010

National Capital Region

Projective Clustering Ensembles

Speaker: Prof. Carlotta Domeniconi
Department of Computer Science
George Mason University

Friday, March 26, 2010
1:00PM- 2:00PM, NVC 325

Abstract

Recent advances in clustering have focused on clustering ensembles and projective clustering approaches. Clustering ensembles offer a solution to challenges inherent to clustering arising from its ill-posed nature: they can provide more robust and stable solutions by making use of the consensus across multiple clustering results, while averaging out emergent spurious structures that arise due to the various biases to which each participating algorithm is tuned, or to the variance induced by different data samples. An orthogonal issue related to clustering is high dimensionality. A common scenario with high-dimensional data is that several clusters may exist in different subspaces comprised of different combinations of features. To capture such local structure of the data, many different projective clustering methods have been proposed.

Despite the advancement in clustering techniques, and their successful applications in real world problems, a major limitation affects the current scenario: clustering ensembles and projective clustering have been developed independently of one another. Thus, the current state of the art lacks a unified framework that is capable of handling both issues: the ill-posed nature of clustering and the curse of dimensionality. To overcome this limitation we propose a new unified framework that can handle both problems. We propose the first formal definition of the problem of projective clustering ensembles, and heuristics to solve it. The objective is to define methods to exploit the information provided by an ensemble of projective clustering solutions to compute a robust consensus projective clustering. The problem is formulated as a multi-objective optimization problem where the objective functions embed both sides of the ensemble components: the data clusterings and the assignments of features to clusters. Our experimental results on real data sets demonstrate the effectiveness of the proposed methods.

Joint work with Francesco Gullo and Andrea Tagarelli

Biography

Carlotta Domeniconi received a Laurea Degree in computer science from the University of Milan, Italy, in 1992, an M.S. in information and communication technologies from the International Institute for Advanced Scientific Studies, Salerno, Italy, in 1997, and a Ph.D. in computer science from the University of California, Riverside, in 2002.

She is Associate Professor in the Department of Computer Science at George Mason University. Her research interests include machine learning, pattern recognition, data mining, and feature relevance estimation, with applications in text mining and bioinformatics. Her research is in part supported by an NSF CAREER Award and a grant from the U.S. Army.