The main focus of the dissertation research is to design effective learning techniques for information retrieval and mining in high-dimensional databases. There are two main aspects in the retrieval and mining research: accuracy and efficiency. The accuracy problem is how to return results which can better match the ground truth, and the efficiency problem is how to evaluate users’ requests and execute learning algorithms as fast as possible. However, these problems are non-trivial because of the complexity of the high-level semantic concepts, the heterogeneous natures of the feature space, the high dimensionality of data representations, the size of the databases. The dissertation is dedicated to addressing these issues. Specifically, this study has five main contributions as follows.

The first contribution is to design a novel manifold learning algorithm, which defines a salient embedding of the high-dimensional data space with little distortion to the original local and global structures. The second contribution is an effective semi-supervised clustering algorithm, which leverages instance-level constraints to achieve better data grouping. The third contribution is a unified framework for partition-based dimension reduction techniques and the efficient solution to determine the optimal dimension partition for a dataset. The fourth contribution studies the similarity search problem in the point-set data, in which meaningful distance functions and lower bounding approximation algorithms are developed. The last contribution is a novel scheme to convert the multiple instance learning problem into the standard supervised learning problem by systematically generating a single feature vector for a bag of instances.

Major: Computer Science

Educational Career:
Bachelor's of Electronic Information Science and Technology, BS, 2004, University of Science and Technology of China
Master's of Computer Science, MS, 2008, University of Central Florida

Committee in Charge:
Kien A. Hua, Chair, Computer Science
Michael Georgiopoulos, Electrical and Computer Engineering
Charles E. Hughes, Computer Science
Sheau-Dong Lang, Computer Science
Xin Li, Mathematics

Approved for distribution by Kien A. Hua, Committee Chair, on October 5, 2009.

The public is welcome to attend.