

Learning Embeddings for Similarity-Based Retrieval

Dr. Vassilis Athitsos,

Boston University

<http://cs-people.bu.edu/athitsos/>

Tuesday, March 13, 2007

2:30-3:45 PM

TECH Center 111

Abstract:

Similarity-based retrieval is the task of identifying database patterns that are the most similar to a query pattern. Retrieving similar patterns is a necessary component of many practical applications, in fields as diverse as computer vision, speech recognition, and bioinformatics. This talk presents BoostMap, a method for efficient similarity-based retrieval in spaces with computationally expensive distance measures. Our method constructs embeddings that map database and query patterns into a vector space with a computationally efficient distance measure. Using such a mapping, similar patterns can be retrieved efficiently - often orders of magnitude faster compared to retrieval using the original distance measure. In the BoostMap method, embedding construction is treated as a machine learning problem, and embedding quality is optimized using information from training data. A key property of the learning-based formulation is that the optimization criterion does not depend on geometric properties and is equally valid in both metric and non-metric spaces. In experiments with several datasets, our method compares favorably to alternative methods for efficient retrieval, and provides highly competitive results for applications such as handwritten character recognition and time series indexing.

Biography:

Dr. Athitsos received the BS degree in mathematics from the University of Chicago in 1995, the MS degree in computer science from the University of Chicago in 1997, and the PhD degree in computer science from Boston University in 2006. In 2005-2006 he worked as a researcher at Siemens Corporate Research, developing methods for database-guided medical image analysis. Since October 2006 he is a postdoctoral research associate at the Computer Science department at Boston University. His research interests include computer vision, machine learning, and data mining. His recent work has focused on efficient similarity-based retrieval, gesture recognition, shape modeling and detection, and medical image analysis.