

EINLADUNG

Zeit: Dienstag, 13. September 2011, 16.00 Uhr

Ort: Room 5056, Ahornstr. 55

Referent: Prof. Michael E. Houle
National Institute for Informatics, Tokyo, Japan

Titel: Intrinsic Dimensionality and its Applications to
Databases and Data Mining

Abstract:

In recent years, important applications in the areas of machine learning, data mining, and bioinformatics have been developed that rely on the search for similarities among data structures as a fundamental operation. Typically, the computation of structural similarity is performed using a computationally-expensive exhaustive search over all possibilities. Effective indexing methods for structural similarity search, together with appropriate measures of inter-structure similarity, could potentially have a significant impact on the overall performance of such applications.

In this presentation, we propose a general framework within which the effectiveness of similarity search and other data operations can be assessed, independently of the nature of the data being considered. Within this framework, we discuss how a form of intrinsic dimensionality in the neighborhood of a test or query item, the "stereological dimension", can be assessed using two measurements of the distance and ranks of items within the neighborhood. Examples will be shown of the use of stereological dimension in the design and performance analysis of a variety of applications, such as similarity search, aggregation of search results, and anomaly detection (a paper describing the latter application received the "Best Research Paper Award" at IEEE ICDM 2010). The presentation will conclude with a discussion of the potential of this framework for the development of effective applications based on structural similarity search.