

EINLADUNG

Zeit: Donnerstag, 15. März 2012, 15:00 Uhr

Ort: Raum 5056, Ahornstr. 55

Referent: Dipl.-Inform. Stephan Günnemann

Titel: Subspace Clustering for Complex Data

Abstract:

The increasing potential of storage technologies and information systems has opened the possibility to conveniently and affordably gather large amounts of complex data. Going beyond simple descriptions of objects by some few characteristics, such data sources range from high dimensional vector spaces over imperfect data to network data describing relations between the objects.

Data Mining is the task of automatically extracting previously unknown and useful patterns from such data. One of the most prominent mining tasks is *clustering*, which aims at grouping similar objects while separating dissimilar ones. Since in today's applications usually many characteristics for each object are recorded, one cannot expect to find similar objects by considering all attributes together. As a general solution to this problem, the paradigm of *subspace clustering* has been introduced, which aims at detecting clusters hidden in subspace projections of the data.

In this talk, we discuss novel methods for effective subspace clustering on various types of complex data. Our methods tackle major open challenges for clustering in subspace projections. We study the problem of redundancy in subspace clustering results and propose models whose solutions contain only non-redundant and, thus, valuable clusters. Since different subspace projections represent different views on the data, often several groupings of the objects are reasonable. Thus, we propose techniques that are not restricted to a single partitioning of the objects but that enable the detection of multiple clustering solutions. Besides tackling these challenges of subspace clustering for the case of vector data, we study the task of subspace clustering on two further data types: imperfect data and network data. We propose integrated mining techniques directly handling errors in the data and simultaneously mining different information sources. Overall, for the first time, meaningful subspace clustering results can be obtained for these types of complex data.

Es laden ein: Die Dozenten der Informatik