CARL VON OSSIETZKY universität OLDENBURG

Fakultät II *Informatik, Wirtschafts- und Rechtswissenschaften*
Department für Informatik

Informatik Kolloquium

Am Donnerstag, dem **16. Juni 2011,** um **16:15** Uhr hält

Dr. Michał Grabowski
Warsaw School of Computer Science

einen Vortrag mit dem Titel

Metrics and similarities in modeling dependencies between continuous and nominal data

Der Vortrag findet im OFFIS, Escherweg 2, Konferenzraum F02 statt.

Abstract:

In its analytical paradigm, classification theory investigates continuous data only. When we are faced with a mixture of continuous and nominal attributes in data records, difficulties are emerging. Usually, the analytical paradigm is maintained by treating nominal attributes as continuous ones via numerical coding of nominal values (often a bit ad hoc). We propose a way of keeping nominal values within the analytical paradigm with no pretending that nominal values are continuous.
The core idea is that the information hidden in nominal values influences the metrics (or the similarity function) on records of continuous and nominal data. An adaptation via genetic algorithms finds the relevant parameters governing the influence of nominal values on the metrics between data records. Our approach works well for classifier induction algorithms where the metrics or similarity is generic, for instance the k-nearest neighbor algorithm or, proposed here, the support of decision tree induction by a similarity function between data. The k-nearest neighbor algorithm working with continuous and nominal data behaves considerably better when nominal values are processed in accordance with our approach. It appears that the idea has some significance. We propose also a generalization of the k-means clustering algorithm to mixtures of continuous and nominal data. No numerical coding of nominal values is involved.
Algorithms of the analytical paradigm using linear and probability machinery, like discriminant adaptive nearest-neighbor or Fisher's linear discriminant analysis, cause problems nevertheless we sketch some possible ways to overcome these problems. Our work on such generalizations is at the very beginning stage yet.

Eingeladen von: Prof. Dr. Ernst-Rüdiger Olderog

Weitere Kolloquiumstermine sind im WWW abrufbar.