RadViz and its application in Data Mining

Lenka Nováková, Olga Štěpánková
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague

Data-mining (DM) is a typical interdisciplinary activity, which requires close cooperation between two types of experts: those from the target domain (e.g. medicine, banking or marketing) and those from the DM or machine learning community. The aim of both these groups of experts is to identify interesting patterns hidden (but implicitly present) in the studied data.

Often, there is a number of different approaches which can be applied to given data as well as number of questions which can be asked during the data mining process and the domain experts have crucial role in making the appropriate choice. They are supposed to help in focusing attention towards the most promising direction while taking into account the domain knowledge and the results, which have been obtained already.

It seems to be most natural to use data visualization for that purpose because human mind excels in prompt interpretation of visual information. It is not surprising that Soukup and Davidson claim that "Visualization is a key in assisting business and data analysis to discover new patterns and trends in their business data sets."

The DM is rarely concerned with data of such simplicity and what is needed is visualization means suited for multidimensional data. One of such methods is offered by the RadViz algorithm that has been designed to map data described by 3 or more attributes (multidimensional data) into a planar picture. This contribution tries to answer the question whether RadViz algorithm can serve the purpose of typical data mining task namely identification of clusters in multidimensional space, visualization trends and detection outliers.