Rough Sets for Data Mining and Knowledge Discovery

Jan Komorowski¹, Lech Polkowski², Andrzej Skowron³

¹ Dept of Computer and Information Science
Norwegian University of Science and Technology
Trondheim, Norway

² Dept. of Mathematics
Warsaw Technical University
Warsaw, Poland

³ Dept of Mathematics, Informatics and Theoretical Mechanics
Warsaw University
Warsaw, Poland

Abstract

The problem of handling imperfect and approximate knowledge has been recently recognized as the crucial issue in solving several complex real-life problems, also in the case of data mining and knowledge discovery. Among many approaches to imperfect knowledge, such as fuzzy sets of Zadeh (1965) various forms of neural networks and evidence theory, to name a few, the theory of information systems and rough sets introduced by Z. Pawlak in 1982 has lately gain substantial attention. Rough sets are attractive from the computational point of view because the underlying concept of an information system is basically a relation in a database. Rough sets also have a rather intuitive common-sense interpretation (i.e. belief and plausibility) having at the same time a sound mathematical definition (i.e. lower and upper bounds).

Several problems concerning data analysis presented in the form of an information system can be solved using the rough set theory. This methodology has been successfully applied in medical data analysis, finance, voice recognition, image processing, process modelling and identification, conflict resolution, etc. It has many advantages such as: efficient algorithms for finding hidden patterns in data, finding minimal sets of data (data reduction), evaluation of the significance of data, generation of relevant sets of decision rules or features from data, straightforward interpretation of results presented as rules, and others.

This tutorial introduces basic rough sets for data and continues with advanced methods for synthesizing approximate classification and decision rules; it then gives methods for dealing with large data sets. The participants will be given an opportunity to gain hands-on experience using Rosetta – our PC-based tool-kit for data analysis with rough sets.

(<http://www.idi.ntnu.no/aleks/rosetta/rosetta.html>)