GROBIAN
An Engine for Rough Information Analysis

Günther Gediga
Fachbereich Psychologie/Methodik
Universität Regensburg
49076 Regensburg, Germany
gediga@psych.fei-univ-regensburg.de

Ivo Dintsch
School of Information and Software Engineering
University of Ulster
Newtownabbey, BT37 OQB, N.Ireland
dintsch@ulst.ac.uk

May 7, 1999

Introduction and Background

GROBIAN is an implementation of the rough information analysis. ROUGH-MAY described detail in [7].
The acronym GROBIAN comes from the German "GROBengen Informations Analyse", which is the
same as "ROUGH set Information Analyse" and the meaning of the German GROBIAN and the English
ROUGH-MAY is (roughly) the same. GROBIAN is bilingual (English/German).

The basis for ROUGH-MAY is rough set data analysis (RSDA) which has recently gained importance in data
mining [10, 11]. ROUGH-MAY enhances RSDA by providing well founded statistical and information theoretic
procedures which turn RSDA into a visible tool for feature selection, data filtering, and prediction by
using only the data at hand without the restrictive additional model assumptions required by most statistical
methods, but also by soft computing procedures. Nevertheless, tests have shown that the prediction quality
of ROUGH-MAY is comparable to standard machine learning or statistical procedures [1, 6].

Functionality

Beside the standard procedures of Rough Set Data Analysis, such as reduce analysis, \( \gamma \) and \( \alpha \) statistics,
and rule generation, GROBIAN has the following enhanced ROUGH-MAY features:

- A randomization test for rule significance including a sequential testing procedure which dramatically
  reduces the computation time [9],
- A structural method for data filtering which may increase the significance of rules [5],
- Model selection based on information theoretic entropy, [6]
- Test-knife validation,
- Training-Testing validation.

A menu driven recording and restriction procedure for data manipulation is an additional feature of GROBIAN.
Developers

1. Main developer: G. Gediga, FB Psychologie und Institut für semantische Informationsverarbeitung, Universität Konstanz, gediga@psycho.uni-konstanz.de

2. I. Dünach, Faculty of Informatics, University of Ulster, i.duenach@ulst.ac.uk

Publications

Publications about the system are [2, 3, 4, 8].

History

- V 0.01 - 0.06: Version for internal use only.
- V 0.07: First published GROMAX version, including significance testing, rough filtering, rough entropy (July 1996).
- V 0.08 - 0.13: Several minor changes and bug removal.
- V 0.14: Removed the RSL library from most parts of GROMAX.
- V 0.15: Added TEST - X-MORE validation procedure.
- V 0.16: Added batch processing from a list of datasets.
- V 0.17: Sequential Randomization Test is now available for significance testing. Added TRAINING SET - TESTING SET validation procedure.
- V 0.18: Real-time windows can be handled as objects (Oct 1997). Added search for attribute sets with minimal rough entropy.

Input

One table which may be partitioned by data restrictions to enable many tables-applications. ASCII-files and specific GROMAX files are possible input formats. There is no direct interaction with DBMS systems (except via ASCII export - import).

Output

The results of the chosen procedure on screen and as an ASCII file.

System requirements

- Processor: Intel 80386 or later,
- MS Windows family (31, Win95, NT),
- 8 MB RAM,
- 1 MB disk space + space for data files.

Data Visualisation

None at present.
Documentation

Some help is available.

Availability

The present β - version is available freely from www.infj.ULATEC.ac.uk/~creez33/grobian/
grobian.html. Since April 1997, it has been downloaded from approximately 400 different sites worldwide.

References


