# GROBIAN
# An Engine for Rough Information Analysis

Günther Gediga
Fachbereich Psychologie/Methodenlehre
Universität Osnabrück
49069 Osnabrück, Germany
gg@Luce.Psycho.Uni-Osnabrueck.DE

Ivo Düntsch
School of Information and Software Engineering
University of Ulster
Newtownabbey, BT 37 0QB, N.Ireland
I.Duentsch@ulst.ac.uk

May 7, 1999

## Introduction and Background

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

The basis for ROUGHIAN is rough set data analysis (RSDA) which has recently gained importance in data mining [10, 11]. ROUGHIAN enhances RSDA by providing well founded statistical and information theoretic procedures which turn RSDA into a viable 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 ROUGHIAN is comparable to standard machine learning or statistical procedures [1, 6].

## Functionality

Beside the standard procedures of Rough Set Data Analysis, such as redust analysis, $\gamma$ and $\alpha$ – statistics, and rule generation, GROBIAN has the following enhanced ROUGHIAN features:

- A randomization test for rule significance including a sequential testing prodecure 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]
- Jack–knife validation,
- Training/Testing validation.

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

## Developers

1. Main developer: G. Gediga, FB Psychologie and Institut für semantische Informationsverarbeitung, Universität Osnabrück, gg@Luce.psycho.uni-osnabrueck.de
2. I. Düntsch, Faculty of Informatics, University of Ulster, I.Duentsch@ulst.ac.uk

### Publications

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

## History

| | |
|---|---|
| V 0.01 – 0.06 | Versions for internal use only. |
| V 0.07 | First published GROBIAN 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 GROBIAN. |
| V 0.15 | Added JACK–KNIFE validation procedure. |
| V 0.16 | Added batch processing from a list of redicts. |
| V 0.17 | Sequential Randomization Test is now available for significance testing. Added TRAINING SET - TESTING SET validation procedure. |
| V 0.18 | Result windows can be handled as objects (Oct 1997). Added search for attribute sets with minimal rough entropy. |

## Input

One table which may be partioned by data restrictions to enable many tables-applications. ASCII-files and specific GROBIAN 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 (3.1, Win9x, 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 $\beta$ – version is available freely from `www.infj.ulst.ac.uk/~cccz23/grobian/grobian.html`. Since April 1997, it has been downloaded from approximately 400 different sites worldwide.

## References


[1] Browne, C., Düntsch, I. & Gediga, G. (1998). IRIS revisited: A comparison of discriminant and enhanced rough set data analysis. In [12], 345–368.

[2] Düntsch, I. & Gediga, G. (1997). The rough set engine GROBIAN. In A. Sydow (Ed.), *Proc. 15th IMACS World Congress, Berlin*, vol 4, 613–618, Berlin. Wissenschaft und Technik Verlag.

[3] Düntsch, I. & Gediga, G. (1998a). GROBIAN. In [12], 555–557.

[4] Düntsch, I. & Gediga, G. (1998b). GROBIAN – An implementation of rough information analysis. *Bulletin of the International Rough Set Society*.

[5] Düntsch, I. & Gediga, G. (1998c). Simple data filtering in rough set systems. *International Journal of Approximate Reasoning*, **18**, 93–106.

[6] Düntsch, I. & Gediga, G. (1998d). Uncertainty measures of rough set prediction. *Artificial Intelligence*, **106**, 77–107.

[7] Düntsch, I. & Gediga, G. (1999). ROUGHIAN – Rough Information Analysis. *International Journal of Intelligent Systems*. To appear.

[8] Düntsch, I., Gediga, G. & Jüttner, A. (1996). GROBIAN – An engine for rough set data analysis. In *Proceedings of the First International Conference on Practical Aspects of Knowledge Management*, Basel.

[9] Gediga, G. & Düntsch, I. (1998). Statistical tools for rule based data analysis. In J. Komorowski, I. Düntsch & A. Skowron (Eds.), *Workshop on "Synthesis Of Intelligent Agent Systems From Experimental Data", ECAI 98*.

[10] Lin, T. Y. & Cercone, N. (Eds.) (1997). Rough sets and data mining, Dordrecht. Kluwer.

[11] Pawlak, Z., Grzymala-Busse, J. W., Słowiński, R. & Ziarko, W. (1995). Rough sets. *Comm. ACM*, **38**, 89–95.

[12] Polkowski, L. & Skowron, A. (Eds.) (1998). Rough sets in knowledge discovery, vol. 2. Heidelberg: Physica–Verlag.